

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:31:02 ; Search time 35.5 seconds  
(without alignments)  
35.769 Million cell updates/sec

Title: US-10-064-903-1

Perfect score: 29

Sequence: 1 RXXHXXH 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	79.3	30	AA013629	Human polypeptide
2	23	79.3	60	ABP10050	Human ORFX protein
3	23	79.3	66	ABP33078	Human ORF2051 prot
4	23	79.3	68	ABP04899	Human ORFX protein
5	23	79.3	80	AAU60888	Propionibacterium
6	23	79.3	92	AAU11260	Human polypeptide
7	23	79.3	115	ABP03095	Human ORFX protein
8	23	79.3	116	ABP01907	Human ORFX protein
9	23	79.3	119	AAW25686	Human protein sequ

10	23	79.3	132	22	AA063312	Human breast cance
11	23	79.3	134	22	AAU67989	Propionibacterium
12	23	79.3	137	22	ABP02968	Human ORFX protein
13	23	79.3	141	22	AAO10851	Human polypeptide
14	23	79.3	167	22	ABP70944	Drosophila melanog
15	23	79.3	218	23	ABP31393	Human ORF366 prote
16	23	79.3	357	22	AAU64020	Propionibacterium
17	23	79.3	462	22	AAU42927	Propionibacterium
18	23	79.3	466	18	AAW09825	UDP-glucose:thiohy
19	23	79.3	508	22	ABW11345	Drosophila melanog
20	23	79.3	543	22	ABG22945	Novel human diagno
21	23	79.3	572	24	ABP77246	N. gonorrhoeae ami
22	23	79.3	635	18	AAW19920	Human Ksr' (kinase
23	23	79.3	880	22	ABG65766	Drosophila melanog
24	23	79.3	1078	24	ABP86069	Human protein kina
25	23	79.3	1133	22	ABG55544	Drosophila melanog
26	23	79.3	1187	22	ABG67666	Drosophila melanog
27	23	79.3	1518	24	ABJ18375	Breast specific re
28	23	79.3	1529	17	AAR97985	CORK potassium cha
29	23	79.3	1575	22	ABG27933	Novel human diagno
30	23	79.3	2424	22	ABG58924	Drosophila melanog
31	23	79.3	3502	22	ABG38382	Drosophila melanog
32	22	75.9	34	21	AAG07703	Arabidopsis thalia
33	22	75.9	37	22	AAW86640	Human immune/haema
34	22	75.9	41	21	AAW34597	Human secreted pro
35	22	75.9	49	22	ABJ17194	Human nervous syst
36	22	75.9	50	23	ABP1516	Human ORF499 prote
37	22	75.9	52	23	ABG93190	S. cerevisiae BAX-
38	22	75.9	53	23	ABP07950	Human ORFX protein
39	22	75.9	55	22	AAU58208	Propionibacterium
40	22	75.9	56	23	ABP32940	Human ORF1913 prot
41	22	75.9	57	22	AAW78747	Human protein SEQ
42	22	75.9	59	22	AAW79731	Human protein SEQ
43	22	75.9	60	23	ABP05469	Human ORFX protein
44	22	75.9	61	22	AAO10714	Human polypeptide
45	22	75.9	61	22	AAG75865	Human colon cancer

#### ALIGNMENTS

#### RESULT 1

AA013629  
ID AA013629 standard; Protein; 30 AA.

XX AC AA013629;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 27521.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX XX WPI; 2001-514838/56.

DR N-PSDB; AAI93560.

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
XX Claim 20; SEQ ID NO 27521; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AAI9941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 30 AA;  
Query Match 79.3%; Score 23; DB 22; Length 30;  
Best Local Similarity 37.5%; Pred. No. 4.3e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 HXXXHXXX 8  
| | | |  
DB 18 HTHHTSH 25

RESULT 2  
ABP10050  
ID ABP10050 standard; Protein; 60 AA.  
XX  
AC ABP10050;  
XX  
DT 25-JUN-2002 (first entry)  
XX  
DE Human ORFX protein sequence SEQ ID NO:20082.  
XX  
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX  
OS Homo sapiens.  
XX  
PN WO200192523-A2.  
XX  
PD 06-DEC-2001.  
XX  
PF 29-MAY-2001; 2001WO-US10836.  
XX  
PR 30-MAY-2000; 2000US-206132P.  
XX  
PR 29-AUG-2000; 2000US-228716P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Leach MD;  
XX  
DR WPI; 2002-106308/14.  
DR N-PSDB; ABN25802.  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders -

XX  
PS Disclosure; SEQ ID 20082; 1037pp; English.  
XX  
CC The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification)). AEN15762 to AEN27282 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 60 AA;  
Query Match 79.3%; Score 23; DB 23; Length 60;  
Best Local Similarity 37.5%; Pred. No. 7.4e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXX 8  
| | | |  
DB 20 HSSHTTH 27

RESULT 3  
ABP33078  
ID ABP33078 standard; Protein; 66 AA.  
XX  
AC ABP33078;  
XX  
DT 09-JUL-2002 (first entry)  
XX  
DE Human ORP2051 protein, SEQ ID NO:4102.  
XX  
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
KW immune modulation; haematopoiesis regulation; tissue growth;  
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
KW behaviour; cancer; proliferative disorder; neurological disorder;  
KW cardiovascular disease; immune system disorder; organ transplantation;  
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neurotropic;  
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
KW dermatological; analgesic; virucide; antibacterial; fungicide.  
XX  
OS Homo sapiens.  
XX  
PN WO200190366-A2.  
XX  
PD 29-NOV-2001.  
XX  
XX 24-MAY-2001; 2001WO-US17076.  
XX  
XX 24-MAY-2000; 2000US-206690P.

XX PA (CURA-) CURAGEN CORP.  
 XX PI Leach MD, Shinkets RA;  
 XX DR WPI: 2002-106200/14.  
 DR N-PSDB; ABN77104.  
 XX  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation -  
 XX  
 PS Claim 10; Page 1282-1283; 2508pp; English.  
 XX  
 CC Sequences ABP31028-ABP3561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals,  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.  
 XX  
 SQ Sequence 66 AA;  
 Query Match 79.3%; Score 23; DB 23; Length 66;  
 Best Local Similarity 37.5%; Pred. No. 8e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 HXXXHXXH 8  
 Db 28 HHTTHTSH 35  
 RESULT 4  
 ABP04899  
 ID ABP04899 standard; Protein; 68 AA.  
 XX  
 AC ABP04899;  
 XX  
 XX 25-JUN-2002 (first entry)  
 DT  
 XX Human ORFX protein sequence SEQ ID NO:9780.  
 DE  
 XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;

KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 OS  
 XX Homo sapiens.  
 XX  
 PN WO200192523-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 PF 29-MAY-2001; 2001WO-US10836.  
 XX  
 PR 30-MAY-2000; 2000US-206132P.  
 PR 29-AUG-2000; 2000US-228716P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 PI Shinkets RA, Leach MD;  
 DR WPI: 2002-106308/14.  
 DR N-PSDB; ABN20651.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders -  
 XX  
 PS Disclosure; SEQ ID 9780; 1037pp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis.  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 68 AA;  
 Query Match 79.3%; Score 23; DB 23; Length 68;  
 Best Local Similarity 37.5%; Pred. No. 8.2e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 HXXXHXXH 8  
 Db 9 HTHSHHTAH 16  
 RESULT 5  
 AAU60888  
 ID AAU60888 standard; Protein; 80 AA.  
 XX  
 AC AAU60888;

```

XX DT 27-FEB-2002 (first entry)
XX DE Propionibacterium acnes immunogenic protein #21784.
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX KW dermatological; osteopathic; neuroprotectant.
XX OS Propionibacterium acnes.
XX PN WO200181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001WO-US12865.
XX PR 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX PR WPI: 2001-616774/71.
XX DR N-PSDE; AAS59613.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX PS Example 1; SEQ ID NO 22083; 1069pp; English.
XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA).
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 80 AA;
XX Query Match 79.3%; Score 23; DB 22; Length 80;
XX Best Local Similarity 37.5%; Pred. No. 9.4e+02;
XX Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX QY 1 HXXXHXXH 8
XX DB 47 HSASHRTH 54
XX RESULT 6
XX AA011260
XX ID AA011260 standard; Protein; 92 AA.
XX AC AA011260;

XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 25152.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US04927.
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX PR WPI: 2001-514838/56.
XX DR N-PSDE; AAI91191.
XX PT Isolated nucleic acids and polypeptides, useful for preventing
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX PT disorders -
XX PS Claim 20; SEQ ID NO 25152; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 92 AA;
XX Query Match 79.3%; Score 23; DB 22; Length 92;
XX Best Local Similarity 37.5%; Pred. No. 1e+03;
XX Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX QY 1 HXXXHXXH 8
XX DB 41 HTTAHGSH 48
XX RESULT 7
XX ABP03095
XX ID ABP03095 standard; Protein; 115 AA.
XX AC ABP03095;
XX DT 24-JUN-2002 (first entry)
XX DE Human ORFX protein sequence SEQ ID NO:6172.
XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX KW degenerative disorder; osteoarthritis; neurodegenerative disorder;

```

KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200192523-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 XX  
 PF 29-MAY-2001; 2001WO-US10836.  
 XX  
 PR 30-MAY-2000; 2000US-206132P.  
 PR 29-AUG-2000; 2000US-228716P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinkets RA, Leach MD;  
 XX  
 DR WPI; 2002-106308/14.  
 DR N-PSDB; ABN18847.  
 XX  
 PS Novel human polypeptides and polynucleotides useful for diagnosing,  
 XX preventing and treating cardiovascular disease, neurodegenerative,  
 XX hyperproliferative disorders and autoimmune disorders -  
 XX  
 PS Disclosure; SEQ ID 6172; 1037pp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypotension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 115 AA;  
 Query Match 79.3%; Score 23; DB 23; Length 115;  
 Best Local Similarity 37.5%; Pred. No. 1.3e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HXXHXXH 8  
 Db 45 HTTHTLH 52  
 RESULT 8  
 ID ABP01907 standard; Protein; 116 AA.  
 XX  
 AC ABP01907;  
 XX

DT 24-JUN-2002 (first entry)  
 XX  
 DE Human ORFX protein sequence SEQ ID NO:3796.  
 XX  
 KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200192523-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 XX  
 PF 29-MAY-2001; 2001WO-US10836.  
 XX  
 PR 30-MAY-2000; 2000US-206132P.  
 PR 29-AUG-2000; 2000US-228716P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinkets RA, Leach MD;  
 XX  
 DR WPI; 2002-106308/14.  
 DR N-PSDB; ABN17659.  
 XX  
 PS Novel human polypeptides and polynucleotides useful for diagnosing,  
 XX preventing and treating cardiovascular disease, neurodegenerative,  
 XX hyperproliferative disorders and autoimmune disorders -  
 XX  
 PS Disclosure; SEQ ID 3796; 1037pp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypotension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 116 AA;  
 Query Match 79.3%; Score 23; DB 23; Length 116;  
 Best Local Similarity 37.5%; Pred. No. 1.3e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HXXHXXH 8  
 Db 46 HSTAHSRH 53



XX AC AAU67989;  
XX DT 27-FEB-2002 (first entry)  
XX DE Propionibacterium acnes immunogenic protein #28895.  
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypotertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX OS Propionibacterium acnes.  
XX FN W0200181581-A2.  
XX PD 01-NOV-2001.  
XX PF 20-APR-2001; 2001WO-US12865.  
XX PR 21-APR-2000; 2000US-139047P.  
XX PR 02-JUN-2000; 2000US-208841P.  
XX PR 07-JUL-2000; 2000US-216747P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX DR WPI: 2001-616774/71.  
XX DR N-PSDB; AAS59785.  
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
PS Example 1; SEQ ID No 29184; 1069pp; English.  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypotertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 134 AA;  
SQ Query Match 79.3%; Score 23; DB 22; Length 134;  
Best Local Similarity 37.5%; Pred. No. 1.4e+03;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 HXXXHXXH 8  
DB 40 HSTASHH 47  
RESULT 12  
ABP02968  
ID ABP02968 standard; Protein; 137 AA.

XX AC ABP02968;  
XX DT 25-JUN-2002 (first entry)  
XX DE Human ORFX protein sequence SEQ ID NO:5918.  
XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX OS Homo sapiens.  
XX PN W0200192523-A2.  
XX PD 06-DEC-2001.  
XX PF 29-MAY-2001; 2001WO-US10836.  
XX PR 30-MAY-2000; 2000US-206132P.  
XX PR 29-AUG-2000; 2000US-228716P.  
XX PA (CURA-) CURAGEN CORP.  
XX PI Shinkets RA, Leach MD;  
XX DR WPI: 2002-106308/14.  
XX DR N-PSDB; ABN18720.  
XX PT Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders -  
PS Disclosure; SEQ ID 5918; 1037pp; English.  
XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN2752 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 137 AA;  
SQ Query Match 79.3%; Score 23; DB 23; Length 137;  
Best Local Similarity 37.5%; Pred. No. 1.4e+03;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 HXXXHXXH 8

```

Db      77 HTHHTTH 84

RESULT 13
AAO10851
ID AAO10851 standard; Protein; 141 AA.
XX
AC AAO10851;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 24743.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
WPI: 2001-514838/56.
XX
N-PSDB; AAI90782.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 20; SEQ ID NO 24743; 1393pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 141 AA;
Query Match 79.3%; Score 23; DB 22; Length 141;
Best Local Similarity 37.5%; Pred. No. 1.5e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
DE 8 HASAHAYH 15
XX

RESULT 14
ABB70944
ID ABB70944 standard; Protein; 167 AA.
XX
Db      77 HTHHTTH 84

RESULT 15
ABP31393
ID ABP31393 standard; Protein; 218 AA.
XX
AC ABP31393;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human ORF366 protein, SEQ ID NO:732.
XX
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibitor; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;

```



KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnary;  
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;  
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.  
 XX Homo sapiens.

OS

XX

XX

PN WO200190366-A2.

XX

XX

PD 29-NOV-2001.

XX

XX

PF 24-MAY-2001; 2001WO-US17076.

XX

XX

PR 24-MAY-2000; 2000US-206690P.

XX

XX

PA (CURA-) CURAGEN CORP.

XX

XX

PI Leach MD, Shimkets RA;

XX

XX

DR WPI; 2002-106200/14.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 86 HTHSHAAH 93

Search completed: November 12, 2003, 09:35:06  
 Job time : 36.5 secs

Novel human polypeptides and polynucleotides useful for diagnosing,  
 preventing and treating cardiovascular disease, neurodegenerative,  
 hyperproliferative disorders and disorders related to organ  
 transplantation

Claim 10; Page 450; 2508pp; English.

Sequences ABP31028-ABP315561 represent 4534 novel human proteins  
 designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 ABN79587 represent cDNAs encoding them. The invention also encompasses  
 polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 polynucleotides, the recombinant production of ORFX proteins, antibodies  
 specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 polypeptides, methods of screening for modulators of ORFX expression or  
 activity, and methods of screening individuals for a predisposition to an  
 ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 range of biological activities, such as cytokine, cell proliferation,  
 cell differentiation, immune modulation, haematopoiesis regulation,  
 tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 chemokinetic activity, haemostatic activity, thrombolytic activity,  
 receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 of bodily characterisitics, fertility and behaviour. ORFX proteins,  
 nucleic acids and antibodies may be used in the treatment of cancers,  
 other proliferative disorders such as psoriasis and benign tumours,  
 neurological disorders such as epilepsy and Alzheimer's disease,  
 cardiovascular diseases, immune system disorders, disorders related to  
 organ transplantation, disorders of tissue growth and regeneration  
 diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 storage disease, and infectious diseases caused by viral, bacterial,  
 fungal and other pathogens. ORFX nucleic acids may also be used as a  
 source of primers and probes, in the detection of ORFX genomic sequences  
 or transcripts, in the identification and cloning of homologous  
 sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 nucleic acids may additionally be used to produce transgenic animals  
 which may be useful for studying the function and/or activity of ORFX  
 protein, and in drug screening. The ORFX proteins may also be used as  
 immunogens to generate specific antibodies, which are useful in the  
 diagnosis, treatment and monitoring of ORFX-associated diseases.

Sequence 218 AA;

Query Match

Best Local Similarity 79.3%; Score 23; DB 23; Length 218;

Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:33:48 ; Search time 14 Seconds  
(without alignments)  
24.178 Million cell updates/sec

Title: US-10-064-903-1  
Perfect score: 29  
Sequence: 1 HXXXHXXH 8

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/PCITUS COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	79.3	139	4	US-09-252-991A-32472
2	23	79.3	511	4	US-09-252-991A-22789
3	23	79.3	573	4	US-09-252-991A-24488
4	23	79.3	635	1	US-08-571-758-10
5	23	79.3	635	1	US-08-909-984A-10
6	23	79.3	635	1	US-08-909-983-10
7	22	75.9	151	2	US-08-858-767-30
8	22	75.9	151	2	US-08-863-028-30
9	22	75.9	179	4	US-09-615-192A-289
10	22	75.9	249	4	US-09-252-991A-22610
11	22	75.9	260	4	US-09-252-991A-20987
12	22	75.9	272	4	US-08-858-207A-447
13	22	75.9	323	4	US-09-328-352-6181
14	22	75.9	355	2	US-08-758-621-4
15	22	75.9	355	3	US-09-107-858-4
16	22	75.9	369	4	US-09-252-991A-25533
17	22	75.9	387	4	US-09-364-230-18
18	22	75.9	388	4	US-09-252-991A-31265
19	22	75.9	412	4	US-09-252-991A-26532
20	22	75.9	413	4	US-09-328-352-5589
21	22	75.9	431	1	US-08-311-023-2
22	22	75.9	447	4	US-09-252-991A-32122
23	22	75.9	481	4	US-09-252-991A-24508
24	22	75.9	504	4	US-09-252-991A-28224
25	22	75.9	516	4	US-09-252-991A-29719
26	22	75.9	533	4	US-09-252-991A-23560
27	22	75.9	542	4	US-09-107-532A-4858

28	22	75.9	559	2	US-08-756-317-7	Sequence 7, Appli
29	22	75.9	559	2	US-08-756-317-10	Sequence 10, Appl
30	22	75.9	559	4	US-09-672-749-2	Sequence 2, Appli
31	22	75.9	559	4	US-09-821-016-1	Sequence 1, Appli
32	22	75.9	582	4	US-09-252-991A-27626	Sequence 27626, A
33	22	75.9	582	4	US-09-252-991A-32678	Sequence 32678, A
34	22	75.9	637	4	US-09-252-991A-28952	Sequence 28952, A
35	22	75.9	706	4	US-09-252-991A-25730	Sequence 25730, A
36	22	75.9	795	4	US-09-193-562D-11	Sequence 11, Appl
37	22	75.9	821	4	US-09-193-562D-12	Sequence 12, Appl
38	22	75.9	834	4	US-09-187-999-11	Sequence 11, Appl
39	22	75.9	872	4	US-08-844-057-2	Sequence 2, Appli
40	22	75.9	872	4	US-09-006-730-2	Sequence 2, Appli
41	22	75.9	876	1	US-08-785-071A-2	Sequence 2, Appli
42	22	75.9	876	3	US-09-012-872-2	Sequence 2, Appli
43	22	75.9	893	4	US-09-328-352-6626	Sequence 6626, Ap
44	22	75.9	897	4	US-09-134-001C-3600	Sequence 3600, Ap
45	22	75.9	905	4	US-09-193-562D-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-252-991A-32472  
; Sequence 32472, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32472  
; LENGTH: 139  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32472

Query Match 79.3%; Score 23; DB 4; Length 139;  
Best Local Similarity 37.5%; Fred. No. 5.7e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 HXXXHXXH 8  
Db 26 HTALHSSH 33

RESULT 2  
US-09-252-991A-22789  
; Sequence 22789, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22789  
; LENGTH: 511  
; TYPE: PRT



QY 1 HXXXHXXH 8  
Db 17 HTSAHTQH 24

RESULT 6  
US-08-909-983-10  
; Sequence 10, Application US/08909983  
; Patent No. 5747288  
; GENERAL INFORMATION:  
; APPLICANT: Rubin, Gerry M.  
; APPLICANT: Therrien, Marc  
; APPLICANT: Chang, Henry C.  
; APPLICANT: Karim, Felix D.  
; APPLICANT: Wasserman, David A.  
; TITLE OF INVENTION: A No. 5747288e1 Protein Kinase Required for Ras  
; TITLE OF INVENTION: Signal Transduction  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/909,983  
; FILING DATE: 12-JUN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/571,758  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A.  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: B96-010  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 635 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
US-08-909-983-10

Query Match 79.3%; Score 23; DB 1; Length 635;  
Best Local Similarity 37.5%; Pred. No. 1.8e+03;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
Db 17 HTSAHTQH 24

RESULT 7  
US-08-858-767-30  
; Sequence 30, Application US/08858767  
; Patent No. 5837468  
; GENERAL INFORMATION:  
; APPLICANT: WANG, Xun  
; APPLICANT: DUVICK, Jonathan P.  
; APPLICANT: BRIGGS, Steven P.  
; TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING  
; TITLE OF INVENTION: METHOD  
; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/858,767  
; FILING DATE: 19-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/481,687  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 33229/325/PIHI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-858-767-30

Query Match 75.9%; Score 22; DB 2; Length 151;  
Best Local Similarity 37.5%; Pred. No. 8.9e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
Db 56 HAFATDH 63

RESULT 8  
US-08-863-028-30  
; Sequence 30, Application US/08863028  
; Patent No. 5853991  
; GENERAL INFORMATION:  
; APPLICANT: WANG, Xun  
; APPLICANT: DUVICK, Jonathan P.  
; APPLICANT: BRIGGS, Steven P.  
; TITLE OF INVENTION: PCR-BASED CDNA SUBTRACTIVE CLONING  
; TITLE OF INVENTION: METHOD  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/863,028  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/858,767

; FILING DATE: 19-MAY-1997  
 ; APPLICATION NUMBER: US 08/481,687  
 ; FILING DATE: 07-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 33229/325/PIHI  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202)672-5300  
 ; TELEFAX: (202)672-5399  
 ; TELEX: 904136  
 ; INFORMATION FOR SEQ ID NO: 30:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 151 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-863-028-30

Query Match 75.9%; Score 22; DB 2; Length 151;  
 Best Local Similarity 37.5%; Pred. No. 8.9e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 56 HAFATDTH 63

## RESULT 9

US-09-615-192A-289  
 ; Sequence 289, Application US/09615192A  
 ; Patent No. 6410718

## GENERAL INFORMATION:

; APPLICANT: Bloksberg, Leonard N.  
 ; APPLICANT: Havukkala, Ilkka  
 ; TITLE OF INVENTION: Materials and Methods for the  
 ; TITLE OF INVENTION: Modification of Plant Lignin Content  
 ; FILE REFERENCE: 11000.1003e4U  
 ; CURRENT APPLICATION NUMBER: US/09/615,192A  
 ; CURRENT FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 08/975,316  
 ; PRIOR FILING DATE: 1997-11-21  
 ; PRIOR APPLICATION NUMBER: US 08/713,000  
 ; PRIOR FILING DATE: 1996-09-11  
 ; PRIOR APPLICATION NUMBER: US 09/169,789  
 ; PRIOR FILING DATE: 1998-10-09  
 ; NUMBER OF SEQ ID NOS: 405  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 289

; LENGTH: 179  
 ; TYPE: PRT

; ORGANISM: Eucalyptus grandis  
 ; US-09-615-192A-289

Query Match 75.9%; Score 22; DB 4; Length 179;  
 Best Local Similarity 37.5%; Pred. No. 1e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 170 HSAHSDH 177

## RESULT 10

US-09-252-991A-22610  
 ; Sequence 22610, Application US/09252991A  
 ; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 22610  
 ; LENGTH: 249  
 ; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-22610

Query Match 75.9%; Score 22; DB 4; Length 249;  
 Best Local Similarity 37.5%; Pred. No. 1.3e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 60 HAAHHHAH 67

## RESULT 11

US-09-252-991A-20987  
 ; Sequence 20987, Application US/09252991A  
 ; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 20987  
 ; LENGTH: 260  
 ; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: (78)

; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.

Query Match 75.9%; Score 22; DB 4; Length 260;  
 Best Local Similarity 37.5%; Pred. No. 1.3e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 113 HLAHRSH 120

## RESULT 12

US-08-858-207A-447  
 ; Sequence 447, Application US/08858207A  
 ; Patent No. 6348328

## GENERAL INFORMATION:

; APPLICANT: Black, Michael  
 ; APPLICANT: Hodgson, John  
 ; APPLICANT: Knowles, David  
 ; APPLICANT: Nicholas, Richard  
 ; APPLICANT: Stodola, Robert  
 ; TITLE OF INVENTION: No. 6348328el Compounds  
 ; NUMBER OF SEQUENCES: 552  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SmithKline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA

; COUNTRY: USA  
 ; ZIP: 19406-0939  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/858,207A  
 ; FILING DATE: 09-MAY-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/017670  
 ; FILING DATE: 14-MAY-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gimmi, Edward R  
 ; REGISTRATION NUMBER: 38,891  
 ; REFERENCE/DOCKET NUMBER: P50475  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 610-270-4478  
 ; TELEFAX: 610-270-5090  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 447:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 272 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: No. 6348328e  
 ; US-08-858-207A-447

Query Match 75.9%; Score 22; DB 4; Length 272;  
 Best Local Similarity 37.5%; Pred. No. 1.4e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 151 HTATHLH 158

RESULT 13  
 US-09-328-352-6181  
 ; Sequence 6181, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 6181  
 ; LENGTH: 323  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 ; US-09-328-352-6181

Query Match 75.9%; Score 22; DB 4; Length 323;  
 Best Local Similarity 37.5%; Pred. No. 1.6e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 303 HIAQHSH 310

RESULT 14  
 US-08-758-621-4  
 ; Sequence 4, Application US/08758621  
 ; Patent No. 5846821  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Guerinet, Mary Lou, and Eide, David J.

; TITLE OF INVENTION: Metal-Regulated Transporters and Uses Therefor  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: LAHIVE & COCKFIELD  
 ; STREET: 60 State Street  
 ; CITY: Boston  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02109-1875  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/758,621  
 ; FILING DATE:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/018,578  
 ; FILING DATE: 29-MAY-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Silveri, Jean M.  
 ; REGISTRATION NUMBER: 39,030  
 ; REFERENCE/DOCKET NUMBER: DCI-099CP  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617)227-7400  
 ; TELEFAX: (617)227-5941  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 355 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-758-621-4

Query Match 75.9%; Score 22; DB 2; Length 355;  
 Best Local Similarity 37.5%; Pred. No. 1.7e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 178 HIHTASH 185

RESULT 15  
 US-09-107-858-4  
 ; Sequence 4, Application US/09107858  
 ; Patent No. 6162900  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Guerinet, Mary Lou et al.  
 ; TITLE OF INVENTION: METAL-REGULATED TRANSPORTERS AND USES THEREFOR  
 ; FILE REFERENCE: DCI-099CPDV  
 ; CURRENT APPLICATION NUMBER: US/09/107,858  
 ; CURRENT FILING DATE: 1998-06-30  
 ; EARLIER APPLICATION NUMBER: 08/758,621  
 ; EARLIER FILING DATE: 1996-11-27  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 355  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; US-09-107-858-4

Query Match 75.9%; Score 22; DB 3; Length 355;  
 Best Local Similarity 37.5%; Pred. No. 1.7e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 178 HIHTASH 185

Wed Nov 12 16:14:07 2003

us-10-064-903-1.ra1

Page 6

Search completed: November 12, 2003, 09:38:08  
Job time : 15 secs

---

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:35:43 ; Search time 23 Seconds  
(without alignments)  
59.739 Million cell updates/sec

Title: US-10-064-903-1

Perfect score: 29

Sequence: 1 HXXXHHX 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/FCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/FCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09D\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US09E\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	79.3	890	15	US-10-156-761-14378
2	22	75.9	61	14	Sequence 14378, A
3	22	75.9	61	15	Sequence 228, App
4	22	75.9	69	15	Sequence 6639, Ap
5	22	75.9	69	9	Sequence 35891, A
6	22	75.9	69	12	US-09-864-761-35891
7	22	75.9	102	12	US-10-029-386-29728
8	22	75.9	102	12	US-10-238-075-517
9	22	75.9	104	10	US-10-231-417-489
10	22	75.9	110	9	US-09-764-864-1330
11	22	75.9	110	9	US-09-864-761-35339
12	22	75.9	114	9	Sequence 46752, A
13	22	75.9	136	10	US-09-864-761-37988
14	22	75.9	161	15	US-09-893-737-74
15	22	75.9	167	9	Sequence 74, Appl
16	22	75.9	179	16	US-09-864-761-34765
17	22	75.9	179	16	Sequence 289, App

16 22 75.9 180 9 US-09-811-284-249 Sequence 249, App  
17 22 75.9 193 9 US-09-191-687B-4 Sequence 4, Appli  
18 22 75.9 193 15 US-10-228-796-4 Sequence 4, Appli  
19 22 75.9 207 9 US-09-804-551B-42 Sequence 42, Appli  
20 22 75.9 221 12 US-10-032-585-7060 Sequence 7060, Ap  
21 22 75.9 340 9 US-09-971-361-10 Sequence 10, Appli  
22 22 75.9 342 15 US-10-156-761-12399 Sequence 12399, A  
23 22 75.9 352 15 US-10-232-563-2 Sequence 2, Appli  
24 22 75.9 359 15 US-10-232-563-6 Sequence 6, Appli  
25 22 75.9 359 15 US-10-232-563-7 Sequence 7, Appli  
26 22 75.9 385 12 US-09-855-612-2 Sequence 2, Appli  
27 22 75.9 385 14 US-10-139-262-2 Sequence 2, Appli  
28 22 75.9 385 15 US-10-255-969-2 Sequence 2, Appli  
29 22 75.9 397 10 US-09-925-300-1531 Sequence 1531, Ap  
30 22 75.9 408 9 US-09-864-761-37954 Sequence 37954, A  
31 22 75.9 420 12 US-10-160-764-84 Sequence 84, Appli  
32 22 75.9 424 15 US-10-156-761-8087 Sequence 8087, Ap  
33 22 75.9 483 10 US-09-905-999-20 Sequence 20, Appli  
34 22 75.9 500 12 US-10-032-585-7530 Sequence 7530, Ap  
35 22 75.9 527 15 US-10-128-714-3378 Sequence 3378, Ap  
36 22 75.9 556 15 US-10-128-714-3561 Sequence 3561, Ap  
37 22 75.9 556 15 US-10-128-714-8378 Sequence 8378, Ap  
38 22 75.9 556 15 US-10-128-714-8561 Sequence 8561, Ap  
39 22 75.9 559 9 US-09-821-016-1 Sequence 1, Appli  
40 22 75.9 559 9 US-09-820-952A-1 Sequence 1, Appli  
41 22 75.9 559 9 US-09-820-721A-1 Sequence 1, Appli  
42 22 75.9 559 10 US-09-364-847-21 Sequence 21, Appli  
43 22 75.9 559 15 US-10-218-519-1 Sequence 1, Appli  
44 22 75.9 559 15 US-10-259-632-1 Sequence 1, Appli  
45 22 75.9 559 15 US-10-266-787-1 Sequence 1, Appli

#### ALIGNMENTS

##### RESULT 1

US-10-156-761-14378  
; Sequence 14378, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 14378  
; LENGTH: 890  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-14378

Query Match 79.3%; Score 23; DB 15; Length 890;  
Best Local Similarity 37.5%; Pred. No. 5.8e+03;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHHX 8

Db 573 HSATHLTH 580

##### RESULT 2

US-10-001-835-228



; Sequence 228, Application US/10001835  
 ; Publication No. US20020160387A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Salceda, Susana  
 ; APPLICANT: Macina, Roberto  
 ; APPLICANT: Recipon, Hervé  
 ; APPLICANT: Cafferey, Robert  
 ; APPLICANT: Sun, Yongming  
 ; APPLICANT: Liu, Chenghua  
 ; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro  
 ; FILE REFERENCE: DEX-0277  
 ; CURRENT APPLICATION NUMBER: US/10/001,835  
 ; CURRENT FILING DATE: 2001-11-20  
 ; PRIOR APPLICATION NUMBER: 60/249,997  
 ; PRIOR FILING DATE: 2000-11-20  
 ; NUMBER OF SEQ ID NOS: 228  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 228  
 ; LENGTH: 61  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapien  
 ; US-10-001-835-228

Query Match 75.9%; Score 22; DB 14; Length 61;  
 Best Local Similarity 37.5%; Pred. No. 1.3e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8

Db 22 HRSTHQAH 29

## RESULT 3

; Sequence 6639, Application US/10106698  
 ; Publication No. US20030109690A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruben et al.  
 ; FILE REFERENCE: PAA0591  
 ; CURRENT APPLICATION NUMBER: US/10/106,698  
 ; CURRENT FILING DATE: 2002-03-27  
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524  
 ; PRIOR FILING DATE: 2000-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/157,137  
 ; PRIOR FILING DATE: 1999-09-29  
 ; PRIOR APPLICATION NUMBER: US 60/163,280  
 ; PRIOR FILING DATE: 1999-11-03  
 ; NUMBER OF SEQ ID NOS: 8564  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 6639  
 ; LENGTH: 61  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (24)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (28)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (49)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (53)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (61)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; US-10-106-698-6639

Query Match

75.9%; Score 22; DB 15; Length 61;

Best Local Similarity 37.5%; Pred. No. 1.3e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8

Db 41 HASDHFAH 48

## RESULT 4

; US-09-864-761-35891  
 ; Sequence 35891, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharron G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Aecmica-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,697  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 35891  
 ; LENGTH: 69  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC009743.1  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3  
 ; OTHER INFORMATION: EST HUMAN HIT: AA641863.1, EVALUE 7.10e-01  
 ; OTHER INFORMATION: SWISSPROT HIT: P04929, EVALUE 3.30e+00  
 ; US-09-864-761-35891



```
; Sequence 35339, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Asemlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35339
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF000507.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 6.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: Q30201, EVALUE 9.00e-16
; OTHER INFORMATION: EST_HUMAN HIT: BE877225.1, EVALUE 8.00e-54
US-09-864-761-35339
Query Match 75.9%; Score 22; DB 9; Length 110;
Best Local Similarity 37.5%; Pred. No. 1.9e-03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

QY 1 HXXXHHXXH 8  
| | | | |  
Db 90 HSGNHSTH 97

RESULT 10  
US-09-864-761-46752  
; Sequence 46752, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Asemlca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 46752  
; LENGTH: 110  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC007389.2  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.84  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
; OTHER INFORMATION: EST\_HUMAN HIT: AA305279.1, EVALUE 4.00e-23

```
; OTHER INFORMATION: SWISSPROT HIT: Q26609, EVALUE 8.00e+00
US-09-864-761-46752

Query Match
Best Local Similarity 75.9%; Score 22; DB 9; Length 110;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
DB 55 HQATHSRH 62

RESULT 11
US-09-864-761-37988
; Sequence 37988, Application US/09864761
; Patent No. US20020049763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006655
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37988
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO DB4394.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.9
```

```
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.99
; OTHER INFORMATION: EST HUMAN HIT: BE743982.1, EVALUE 8.00e-54
; OTHER INFORMATION: SWISSPROT HIT: Q30201, EVALUE 9.00e-16
US-09-864-761-37988
```

```
Query Match
Best Local Similarity 75.9%; Score 22; DB 9; Length 114;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
DB 94 HSGNHSTH 101
```

```
RESULT 12
US-09-893-737-74
; Sequence 74, Application US/09893737
; Patent No. US20020110855A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/09/893,737
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-74
```

```
Query Match
Best Local Similarity 75.9%; Score 22; DB 10; Length 136;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8
DB 109 HGTAAH 116
```

```
RESULT 13
US-10-213-880-4
; Sequence 4, Application US/10213880
; Publication No. US2003008083A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Rasco-Gaunt, Sonriza
; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Metal-Binding Proteins
; FILE REFERENCE: BB15.3 US NA
; CURRENT APPLICATION NUMBER: US/10/213,880
; CURRENT FILING DATE: 2002-08-07
; PRIOR APPLICATION NUMBER: U.S. 60/310,522
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Momordica charantia
US-10-213-880-4
```

```
Query Match
75.9%; Score 22; DB 15; Length 161;
```

```
Best Local Similarity 37.5%; Pred. No. 2.5e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8
Db 15 HSHSHSGH 22

RESULT 14
US-09-864-761-34765
; Sequence 34765, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34765
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ACC06371.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
```

```
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EST_HUMAN HIT: BE958003.1, EVALUE 4.60e+00
US-09-864-761-34765
```

```
Query Match 75.9%; Score 22; DB 9; Length 167;
Best Local Similarity 37.5%; Pred. No. 2.6e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy 1 HXXXHXXH 8
Db 112 HMHTHTSH 119
```

```
RESULT 15
US-10-174-693-289
; Sequence 289, Application US/10174693
; Publication No. US20030131373A1
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003C5
; CURRENT APPLICATION NUMBER: US/10/174,693
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 09/615,192
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 407
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 289
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-10-174-693-289
```

```
Query Match 75.9%; Score 22; DB 16; Length 179;
Best Local Similarity 37.5%; Pred. No. 2.7e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy 1 HXXXHXXH 8
Db 170 HSHSHSGH 177
```

```
Search completed: November 12, 2003, 09:44:49
Job time : 23 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:33:02 ; Search time 13.5 Seconds  
(without alignments)  
56.989 Million cell updates/sec

Title: US-10-064-903-1

Perfect score: 29

Sequence: 1 HXXVHXXH 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616882 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.\*

2: Pir2.\*

3: Pir3.\*

4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	23	79.3	152	2	C72662
2	23	79.3	177	2	T26468
3	23	79.3	240	2	F82790
4	23	79.3	327	2	AC2120
5	23	79.3	342	2	T15850
6	23	79.3	382	2	T35709
7	23	79.3	472	2	T27755
8	23	79.3	508	2	S59870
9	23	79.3	510	2	S51124
10	23	79.3	606	2	B69805
11	23	79.3	826	2	T46060
12	23	79.3	826	2	T46061
13	22	75.9	52	2	S63324
14	22	75.9	61	2	AC0287
15	22	75.9	121	2	D82711
16	22	75.9	135	2	I49275
17	22	75.9	144	2	H75636
18	22	75.9	177	2	S85780
19	22	75.9	198	2	B83717
20	22	75.9	208	2	T35454
21	22	75.9	237	2	S19103
22	22	75.9	263	2	G75590
23	22	75.9	306	2	I49068
24	22	75.9	312	2	T27004
25	22	75.9	316	2	D71375
26	22	75.9	325	2	T44782
27	22	75.9	337	1	A42654
28	22	75.9	339	1	S45605
29	22	75.9	339	1	S47643

## ALIGNMENTS

### RESULT 1

C72662  
hypothetical protein APE0723 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: C72662

R.Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takaiawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Taraka, T.; Kudoh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix  
A:Reference number: A72450; PMID:99310339; PMID:10382966

A:Accession: C72662

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-152 <KAW>

A:Cross-references: DBJ:AP000060; NID:G5104188; PIDN:BAA75699.1; PID:d1043485; PID:G51

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE0723

Query Match 79.3%; Score 23; DB 2; Length 152;  
Best Local Similarity 37.5%; Pred. No. 3e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY	1	HXXVHXXH	8
DB	15	HSTTHAAH	22

### RESULT 2

T26468  
hypothetical protein Y11D7A.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T26468

R:Steward, C.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20218

A:Accession: T26468

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-177 <WIL>

A:Cross-references: EMBL:AL032632; PIDN:CAA21589.1; GSPDB:GN000022; CESP:Y11D7A.1

A:Experimental source: clone Y11D7A

C:Genetics:

A:Gene: CESP:Y11D7A.1

A:Map position: 4

A:Introns: 48/1; 102/1; 128/1

Query Match 79.3%; Score 23; DB 2; Length 177;  
Best Local Similarity 37.5%; Pred. No. 3.4e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

30	22	75.9	340	2	T37030	alcohol dehydrogen
31	22	75.9	341	2	E83340	hypothetical prote
32	22	75.9	345	2	T16935	hypothetical prote
33	22	75.9	355	2	T52183	zinc transporter 2
34	22	75.9	364	2	JCS800	peptidylglycine mo
35	22	75.9	383	2	A55739	(MIC) protein MHC
36	22	75.9	384	2	T23604	hypothetical prote
37	22	75.9	394	2	E87606	hypothetical prote
38	22	75.9	403	2	C96757	hypothetical prote
39	22	75.9	416	2	A32947	filaggrin precurs
40	22	75.9	419	2	JQ2254	farne-yl-diphospha
41	22	75.9	420	2	G95107	gamma-glutamyl rho
42	22	75.9	420	2	A97976	glutamate-5-semial
43	22	75.9	424	2	T01383	GTPase-activating
44	22	75.9	427	2	I51580	XFRH2 protein - Af
45	22	75.9	440	2	B71293	hypothetical prote

QY 1 HXXXHXXH 8  
 Db 140 HTTVHSSH 147

RESULT 3  
 F82790  
 GMP synthase XF0560 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: F82790  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: F82790  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-240 <SIM>  
 A:Cross-references: GB:AE003849; GB:AE003849; NID:G9105416; PIDN:AAF83370.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Camargo, D.M.; Carrer, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, B.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF0560

Query Match 79.3%; Score 23; DB 2; Length 240;  
 Best Local Similarity 37.5%; Pred. No. 4.5e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 142 HFSAHATH 149

RESULT 4  
 AC2120  
 cytochrome c oxidase chain II [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AC2120  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
 Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AC2120  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-327 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA074213.1; PID:gl7131606; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: coxB  
 C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology  
 C:Keywords: copper; electron transfer; membrane-associated complex; respiratory chain  
 F:214, 249, 253, 260/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted  
 F:249, 251, 253, 257/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted

F:251/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 79.3%; Score 23; DB 2; Length 327;  
 Best Local Similarity 37.5%; Pred. No. 6e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 126 HASAHVAH 133

RESULT 5  
 T15850  
 hypothetical protein C56C10.10 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T15850  
 R:Fulton, L.  
 submitted to the EMBL Data Library, June 1995  
 A:Description: The sequence of C. elegans cosmid C56C10.  
 A:Reference number: Z18417  
 A:Accession: T15850  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-342 <FUL>  
 A:Cross-references: EMBL:U29488; NID:G868238; PID:G868248; PIDN:AAA68778.1; CESP:C56C1  
 A:Experimental source: strain Bristol N2  
 C:Genetics:  
 A:Gene: CESP:C56C10.10  
 A:Introns: 51/2; 144/2; 204/3; 241/3; 295/3

Query Match 79.3%; Score 23; DB 2; Length 342;  
 Best Local Similarity 37.5%; Pred. No. 6.2e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 128 HSHAHATH 135

RESULT 6  
 T35709  
 hypothetical protein SC7H1.14 SC7H1.14 - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
 C:Accession: T35709  
 R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, January 1998  
 A:Reference number: Z21548  
 A:Accession: T35709  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-382 <MR>  
 A:Cross-references: EMBL:AL021411; PIDN:CAA16201.1; GSPDB:GN00070; SCOEDB:SC7H1.14  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCOEDB:SC7H1.14

Query Match 79.3%; Score 23; DB 2; Length 382;  
 Best Local Similarity 37.5%; Pred. No. 6.9e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
 Db 370 HAARHAH 377

RESULT 7  
 T27755  
 hypothetical protein ZK1320.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 23-Dec-2002  
 C:Accession: T27755

R.Berks, M.  
 Submitted to the EMBL Data Library, December 1994  
 A:Reference number: Z20414  
 A:Accession: T27755  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-472 <WIL>  
 A:Cross-references: EMBL:Z46934; PIDN:CAA87047.1; GSPDB:GN00020; CESP:ZK1320.9  
 A:Experimental source: clone ZK1320  
 C:Genetics:  
 A:Gene: CESP:ZK1320.9  
 A:Map position: 2  
 A:Introns: 19/2; 55/1; 106/1; 160/1; 186/1; 323/1; 411/3  
 C:Superfamily: acetyl-CoA hydrolase

Query Match 79.3%; Score 23; DB 2; Length 472;  
 Best Local Similarity 37.5%; Pred. No. 8.9e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HXXXHXXH 8  
 DB 198 HTVHSH 205

RESULT 8  
 S59870  
 fork head domain protein crocodile - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C:Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 21-Jul-2000  
 C:Accession: S59870; A46178  
 R:Hackler, U.; Kaufmann, E.; Hartmann, C.; Juergens, G.; Knoechel, W.; Jaecckle, H.  
 EMBO J. 14, 5306-5317, 1995  
 A:Title: The *Drosophila* fork head domain protein crocodile is required for the establishment of the head.  
 A:Reference number: S59870; PMID:7489720  
 A:Accession: S59870  
 A>Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-508 <HAB>  
 R:Hackler, U.; Grossniklaus, U.; Gehring, W.J.; Jackle, H.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 8754-8758, 1992  
 A:Title: Developmentally regulated *Drosophila* gene family encoding the fork head domain.  
 A:Reference number: A46178; PMID:92409595; PMID:1356269  
 A:Accession: A46178  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 55-182 <HAC>  
 A:Cross-references: GB:M96440; NID:q157425; PIDN:AAF02177.1; PID:g6042185  
 A:Note: sequence extracted from NCBI backbone (NCBIP:114222)  
 C:Genetics:  
 A:Gene: *croc*  
 A:Cross-references: FlyBase:FBgn0014143  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 F:70-161/Domain: fork head DNA-binding domain homology <FHD>

Query Match 79.3%; Score 23; DB 2; Length 508;  
 Best Local Similarity 37.5%; Pred. No. 8.9e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HXXXHXXH 8  
 DB 207 HMAAAH 214

RESULT 9  
 S55124  
 probable membrane protein YMR177w - yeast (*Saccharomyces cerevisiae*)  
 N:Alternate names: hypothetical protein YMR177w  
 C:Species: *Saccharomyces cerevisiae*  
 C:Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 06-Feb-1998  
 C:Accession: S55124  
 R:Churcher, C.M.  
 Submitted to the EMBL Data Library, June 1995  
 A:Reference number: S55118

A:Accession: S55124  
 A:Molecule type: DNA  
 A:Residues: 1-510 <CHU>  
 A:Cross-references: EMBL:Z49809; NID:G854440; PID:G854447; MIPS:YMR177w  
 A:Experimental source: strain AB972  
 C:Genetics:  
 A:Gene: SGD:MMT1  
 A:Cross-references: SGD:S0004789; MIPS:YMR177w  
 A:Map position: 13R  
 C:Keywords: transmembrane protein  
 F:168-184/Domain: transmembrane #status predicted <TM1>  
 F:232-248/Domain: transmembrane #status predicted <TM2>  
 F:332-348/Domain: transmembrane #status predicted <TM3>  
 F:350-366/Domain: transmembrane #status predicted <TM4>

Query Match 79.3%; Score 23; DB 2; Length 510;  
 Best Local Similarity 37.5%; Pred. No. 8.9e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HXXXHXXH 8  
 DB 137 HTHSHAH 144

RESULT 10  
 B69805  
 conserved hypothetical protein yfix - *Bacillus subtilis*  
 C:Species: *Bacillus subtilis*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 C:Accession: B69805  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
 C.; Bron, S.; Broutillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galie  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Lallo, M.F  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
 Y. M.; Ogawa, K.; Ogiwara, K.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetell  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero  
 kreuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyana  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: B69805  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-606 <KUN>  
 A:Cross-references: GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CAB13672.1; PID:el182833  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: *yfix*

Query Match 79.3%; Score 23; DB 2; Length 606;  
 Best Local Similarity 37.5%; Pred. No. 1e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HXXXHXXH 8  
 DB 244 HSTSHITH 251

RESULT 11  
 T46060  
 hypothetical protein T18N14.20 - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T46060  
 R:Deiseny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K.;  
 submitted to the Protein Sequence Database, December 1999  
 A:Reference number: Z23013



A;Accession: T46060  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-826 <DEL>  
 A;Cross-references: EMBL:AL32968  
 A;Experimental source: cultivar Columbia; BAC clone T18N14  
 C;Genetics:  
 A;Map position: 3  
 A;Introns: 476/3; 796/2  
 A;Note: T18N14.20

Query Match 79.3%; Score 23; DB 2; Length 826;  
 Best Local Similarity 37.5%; Pred. No. 1.4e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXHXXH 8  
 Db 446 HTYAKSSH 453

## RESULT 12

hypothetical protein T18N14.30 - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cross)  
 C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C;Accession: T46061  
 R;Deisner, M.; Berger, C.; Cooke, R.; Grellert, F.; Laudie, M.; Mewes, H.W.; Lemcke, K.;  
 submitted to the Protein Sequence Database, December 1999  
 A;Reference number: 223013  
 A;Accession: T46061  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-826 <DEL>  
 A;Cross-references: EMBL:AL32968  
 A;Experimental source: cultivar Columbia; BAC clone T18N14  
 C;Genetics:  
 A;Map position: 3  
 A;Introns: 476/3; 796/2  
 A;Note: T18N14.30

Query Match 79.3%; Score 23; DB 2; Length 826;  
 Best Local Similarity 37.5%; Pred. No. 1.4e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXHXXH 8  
 Db 446 HTYAKSSH 453

## RESULT 13

hypothetical protein YNL338w - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: hypothetical protein N0170  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 27-Apr-1996 #sequence\_revision 03-May-1996 #text\_change 19-Apr-2002  
 C;Accession: S63324  
 R;Obermaier, B.; Piravandi, E.; Rinke, M.  
 submitted to the Protein Sequence Database, April 1996  
 A;Reference number: S63317  
 A;Accession: S63324  
 A;Molecule type: DNA  
 A;Residues: 1-52 <OBE>  
 A;Cross-references: EMBL:Z71614; NID:gl302466; PID:e239576; PID:gl302467; GSPDB:GNC0014;  
 A;Experimental source: strain S288C  
 C;Genetics:  
 A;Gene: MIPS:YNL338w  
 A;Cross-references: SGD:S0005282  
 A;Map position: 14L

Query Match 75.9%; Score 22; DB 2; Length 52;  
 Best Local Similarity 37.5%; Pred. No. 1.7e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXHXXH 8  
 Db 38 HTHTHTHH 45

## RESULT 14

hypothetical protein YP02354 [imported] - Yersinia pestis (strain CO92)  
 C;Species: Yersinia pestis  
 C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
 C;Accession: AC0287  
 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.E.;  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davis, R.M.; Davis, P.; Dougan, G.;  
 Li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,  
 Nature 413, 523-527, 2001  
 A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A;Reference number: AB0001; MUID:21470413; PMID:11586360  
 A;Accession: AC0287  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-61 <KUR>  
 A;Cross-references: GB:AL590842; PIDN:CAC91159.1; PID:gl5980351; GSPDB:GNC00175  
 C;Genetics:  
 A;Gene: YP02354

Query Match 75.9%; Score 22; DB 2; Length 61;  
 Best Local Similarity 37.5%; Pred. No. 2e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXHXXH 8  
 Db 47 HTHTHTSH 54

## RESULT 15

hypothetical protein XF1205 [imported] - Xylella fastidiosa (strain 9a5c)  
 C;Species: Xylella fastidiosa  
 C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C;Accession: D82711  
 R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ  
 Nature 406, 151-157, 2000  
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A;Reference number: A82515; MUID:20365717; PMID:10910347  
 A;Note: for a complete list of authors see reference number A59328 below  
 A;Accession: D82711  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-121 <SIM>  
 A;Cross-references: GB:AE003954; GB:AE003849; NID:99106165; PIDN:AAF84015.1; GSPDB:GNC0  
 A;Experimental source: strain 9a5c  
 R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier,  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A;Authors: Ferreira, J.C.A.; Perro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fro  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Lai  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,  
 A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y  
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawas  
 A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva  
 M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Varjovski-Almeida, S.; Vettore, A.L.;  
 A;Reference number: A59328  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: XF1205

Query Match 75.9%; Score 22; DB 2; Length 121;  
 Best Local Similarity 37.5%; Pred. No. 3.7e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXHXXH 8

Db 63 HPTTTEH 70

Search completed: November 12, 2003, 09:37:28  
Job time : 15.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:31:57 ; Search time 10 seconds

(without alignments)  
37.621 Million cell updates/sec

Title: US-10-064-903-1

Perfect score: 29

Sequence: 1 HXXHXXH 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	79.3	508	1	CROC DROME
2	23	79.3	510	1	YMA3_YEAST
3	23	79.3	880	1	BRC4_DROME
4	23	79.3	890	1	SYA_STRCO
5	23	79.3	1509	1	GSRI_HUMAN
6	22	75.9	52	1	YN78_YEAST
7	22	75.9	316	1	Y034_TREPA
8	22	75.9	337	1	ADH1_BACST
9	22	75.9	339	1	ADH2_BACST
10	22	75.9	339	1	ADH3_BACST
11	22	75.9	416	1	FILIA_HUMAN
12	22	75.9	419	1	PFTB_PEA
13	22	75.9	420	1	PROA_NEIMB
14	22	75.9	420	1	PROA_STRPN
15	22	75.9	427	1	FKH2_YENLA
16	22	75.9	440	1	Y693_TREPA
17	22	75.9	472	1	SLX1_DROME
18	22	75.9	483	1	CLX1_MOUSE
19	22	75.9	539	1	DOP2_DROME
20	22	75.9	559	1	PHAA_PSEOL
21	22	75.9	590	1	SVY3_HUMAN
22	22	75.9	594	1	SYA_BORBU
23	22	75.9	596	1	FRDA_SHEFR
24	22	75.9	605	1	SYA_TREPA
25	22	75.9	679	1	TKT1_YEAST
26	22	75.9	787	1	AGL2_BACTQ
27	22	75.9	842	1	SYA_CAMEJ
28	22	75.9	860	1	SYA_VIECH
29	22	75.9	860	1	SYA_VIBPA
30	22	75.9	860	1	SYA_VIBVU
31	22	75.9	863	1	SYA_THEMA
32	22	75.9	867	1	SYA_AQUAE
33	22	75.9	867	1	SYA_FUSNN

#### RESULT 1

ID	CROC DROME	STANDARD;	PRT;	508 AA.
AC	P32027; Q9VFP32;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Fork head domain protein crocodile (FKH protein FDI).			
GN	CROC OR FD78E OR FDI OR CG5069.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Canton-S;			
RX	MEDLINE=96080166; PubMed=7489720;			
RA	Haecker U., Kaufmann E., Hartmann C., Juergens G., Knoechel W.,			
RA	Jaeckle H.;			
RT	"The Drosophila fork head domain protein crocodile is required for			
RT	the establishment of head structures";			
RL	EMBO J. 14:5306-5317(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Berkely;			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,			
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.N., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,			
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,			
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,			

#### ALIGNMENTS

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*,"  
RL Science 287:2185-2195(2000).  
[3]  
RN SEQUENCE OF 55-182 FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND  
RP DEVELOPMENTAL STAGE.  
RX MEDLINE=92409595; PubMed=1356269;  
RA Haacker U., Grossniklaus U., Gehring W.J., Jaekle H.;  
RT "Developmentally regulated *Drosophila* gene family encoding the fork  
head domain,"  
RL Proc. Natl. Acad. Sci. U.S.A. 89:8754-8758(1992).  
CC -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF HEAD STRUCTURES.  
CC -!- REQUIRED TO FUNCTION AS AN EARLY PATTERNING GENE IN THE ANTERIOR-  
CC MOST BLASTODERM HEAD SEGMENT ANLAGE AND FOR THE ESTABLISHMENT OF A  
CC SPECIFIC HEAD SKELETAL STRUCTURE THAT DERIVES FROM THE NON-  
CC ADJACENT INTERCALARY SEGMENT AT A LATER STAGE OF EMBRYOGENESIS.  
CC HANDS THE CONSENSUS DNA SEQUENCE 5'-(AG)TAAATTC]A-3'.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN EARLY BLASTODERM EMBRYOS IN  
CC ANTERIOR AND POSTERIOR GUT PRECURSORS, AND, LATER IN A SUBSET OF  
CC CELLS IN CENTRAL NERVOUS SYSTEM.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYOGENESIS, MAXIMALLY  
CC DURING THE 5-12 HOUR PERIOD.  
CC -!- SIMILARITY: Contains 1 fork-head domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; S80254; AAB35643.1; -;  
DR EMBL; AE003594; AAF01727.1; -;  
DR EMBL; M96440; AAF02177.1; -;  
DR PIR; S59870; S59870.  
DR HSSP; Q63245; 2HFH.  
DR TRANSFAC; T02291; -;  
DR FlyBase; FBgn0014143; CROC.  
DR GO; GO:0005634; C:nucleus; IDA.  
DR InterPro; IPR001766; TF\_Fork\_head.  
DR Pfam; PF00250; Fork\_head; 1.  
DR PRINTS; PR00053; FORKHEAD.  
DR ProDom; PD000425; TF\_Fork\_head; 1.  
DR SMART; SM00339; FH; 1.  
DR PROSITE; PS00657; FORK HEAD 1; 1.  
DR PROSITE; PS00658; FORK HEAD 2; 1.  
DR PROSITE; PS00039; FORK HEAD 3; 1.  
DR DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation.  
FT DOMAIN 34 40 POLY-ALA.  
FT DNA\_BIND 69 160 FORK-HEAD.  
FT DOMAIN 161 165 POLY-ARG.  
FT DOMAIN 301 304 POLY-ALA.  
FT DOMAIN 377 380 POLY-ASN.  
FT DOMAIN 389 403 POLY-GLY.  
FT DOMAIN 452 461 POLY-ALA.  
FT DOMAIN 466 473 POLY-HIS.  
FT VARIANT 122 122 L -> F (IN ALLELE CROC-75-3).  
FT VARIANT 453 453 A -> V (IN ALLELE CROC-75-3).  
SQ SEQUENCE 508 AA; 54516 MW; 2EPED1D8F63016D6 CRC64;  
  
Query Match 79.3%; Score 23; DB 1; Length 508;  
Best Local Similarity 37.5%; Pred. No. 4.4e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*,"  
RL Science 287:2185-2195(2000).  
[3]  
RN SEQUENCE OF 55-182 FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND  
RP DEVELOPMENTAL STAGE.  
RX MEDLINE=92409595; PubMed=1356269;  
RA Haacker U., Grossniklaus U., Gehring W.J., Jaekle H.;  
RT "Developmentally regulated *Drosophila* gene family encoding the fork  
head domain,"  
RL Proc. Natl. Acad. Sci. U.S.A. 89:8754-8758(1992).  
CC -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF HEAD STRUCTURES.  
CC -!- REQUIRED TO FUNCTION AS AN EARLY PATTERNING GENE IN THE ANTERIOR-  
CC MOST BLASTODERM HEAD SEGMENT ANLAGE AND FOR THE ESTABLISHMENT OF A  
CC SPECIFIC HEAD SKELETAL STRUCTURE THAT DERIVES FROM THE NON-  
CC ADJACENT INTERCALARY SEGMENT AT A LATER STAGE OF EMBRYOGENESIS.  
CC HANDS THE CONSENSUS DNA SEQUENCE 5'-(AG)TAAATTC]A-3'.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN EARLY BLASTODERM EMBRYOS IN  
CC ANTERIOR AND POSTERIOR GUT PRECURSORS, AND, LATER IN A SUBSET OF  
CC CELLS IN CENTRAL NERVOUS SYSTEM.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYOGENESIS, MAXIMALLY  
CC DURING THE 5-12 HOUR PERIOD.  
CC -!- SIMILARITY: Contains 1 fork-head domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; S80254; AAB35643.1; -;  
DR EMBL; AE003594; AAF01727.1; -;  
DR EMBL; M96440; AAF02177.1; -;  
DR PIR; S59870; S59870.  
DR HSSP; Q63245; 2HFH.  
DR TRANSFAC; T02291; -;  
DR FlyBase; FBgn0014143; CROC.  
DR GO; GO:0005634; C:nucleus; IDA.  
DR InterPro; IPR001766; TF\_Fork\_head.  
DR Pfam; PF00250; Fork\_head; 1.  
DR PRINTS; PR00053; FORKHEAD.  
DR ProDom; PD000425; TF\_Fork\_head; 1.  
DR SMART; SM00339; FH; 1.  
DR PROSITE; PS00657; FORK HEAD 1; 1.  
DR PROSITE; PS00658; FORK HEAD 2; 1.  
DR PROSITE; PS00039; FORK HEAD 3; 1.  
DR DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation.  
FT DOMAIN 34 40 POLY-ALA.  
FT DNA\_BIND 69 160 FORK-HEAD.  
FT DOMAIN 161 165 POLY-ARG.  
FT DOMAIN 301 304 POLY-ALA.  
FT DOMAIN 377 380 POLY-ASN.  
FT DOMAIN 389 403 POLY-GLY.  
FT DOMAIN 452 461 POLY-ALA.  
FT DOMAIN 466 473 POLY-HIS.  
FT VARIANT 122 122 L -> F (IN ALLELE CROC-75-3).  
FT VARIANT 453 453 A -> V (IN ALLELE CROC-75-3).  
SQ SEQUENCE 508 AA; 54516 MW; 2EPED1D8F63016D6 CRC64;  
  
Query Match 79.3%; Score 23; DB 1; Length 508;  
Best Local Similarity 37.5%; Pred. No. 4.4e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
Db 207 HMAAHAAH 214  
  
RESULT 2  
YMA43 YEAST STANDARD; PRT; 510 AA.  
AC Q03218;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Hypothetical 56.2 kDa protein in SIP18-SPT21 intergenic region.  
GN YMR177W OR YMR010.07.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX PubMed=9169872;  
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
RA Rice P., Skelton J., Walsh S., Whitehead S., Bartell B.G.;  
RA "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome  
XIII,"  
RT Nature 387:90-93(1997).  
RL Nature 387:90-93(1997).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: STRONG, TO YEAST YPL224C.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Z49808; CAA89910.1; -;  
DR PIR; S55124; S55124.  
DR SGD; S0004789; WMT1.  
DR GO; GO:0005739; C:mitochondrion; IDA.  
DR GO; GO:0006879; P:iron ion homeostasis; IGI.  
DR InterPro; IPR002524; Cation\_efflux.  
DR Pfam; PF01545; Cation\_efflux; 1.  
DR TIGRFAWS; TIGR01297; CDF; 1.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 165 185 POTENTIAL.  
FT TRANSMEM 194 214 POTENTIAL.  
FT TRANSMEM 241 261 POTENTIAL.  
FT TRANSMEM 286 306 POTENTIAL.  
FT TRANSMEM 333 353 POTENTIAL.  
FT TRANSMEM 356 376 POTENTIAL.  
SQ SEQUENCE 510 AA; 56209 MW; F3CC9A230FB5DB87 CRC64;  
  
Query Match 79.3%; Score 23; DB 1; Length 510;  
Best Local Similarity 37.5%; Pred. No. 4.4e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 HXXXHXXH 8  
Db 137 HTHSHAAH 144  
  
RESULT 3  
BRC4 DROME STANDARD; PRT; 880 AA.  
ID BRC4 DROME Q9W575;  
AC Q24206; Q46064; Q9W575;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Broad-complex core-protein isoform 6.  
GN BR OR BR-C OR EG:17A9.1 OR EG:25D2.1 OR EG:123F11.1 OR  
GN CG11491/CG11514.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota, Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ecnephroidea; Drosophilidae; Drosophila.  
OX NCBI TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A., DEVELOPMENTAL STAGE, AND CHARACTERIZATION OF  
RP ISOFORMS.  
RC TISSUE=imaginal disks, and Larva;  
RX MEDLINE=96299417; PubMed=8660872;  
RA Bayer C.A., Holley B., Fristrom J.W.;  
RT "A switch in broad-complex zinc-finger isoform expression is regulated  
RT posttranscriptionally during the metamorphosis of Drosophila imaginal  
RT discs";  
RL Dev. Biol. 177:1-14(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Oregon-R;  
RX MEDLINE=20196011; PubMed=10731137;  
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,  
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,  
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,  
RA Minano B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,  
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo B.,  
RA Molelell J., Peter A., Schaeffler P., Werner M., Mourikioti F.,  
RA Beinert N., Dowe G., Schaefer U., Jaekle H., Buchanan A.,  
RA Callisier D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,  
RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,  
RA Glover D.M.;  
RT "From sequence to chromosome: the tip of the X chromosome of D.  
RT melanogaster";  
RL Science 287:2220-2222(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Paibis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclev J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
RN [4]  
RP CHARACTERIZATION OF ISOFORMS, AND MUTATIONAL ANALYSIS.  
RX MEDLINE=97384928; PubMed=9242423;  
RA Bayer C.A., von Kalm L., Fristrom J.W.;  
RT "Relationships between protein isoforms and genetic functions  
RT demonstrate functional redundancy at the Broad-Complex during  
RT Drosophila metamorphosis";  
RL Dev. Biol. 187:267-282(1997).  
CC -!- FUNCTION: BROAD-COMPLEX PROTEINS ARE REQUIRED FOR PUFFING AND  
CC TRANSCRIPTION OF SALIVARY GLAND LATE GENES DURING METAMORPHOSIS.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=6;  
CC Name=6; Synonyms=BCORE-Z4;  
CC IsoId=Q24206-1; Sequence=Displayed;  
CC Name=1; Synonyms=BCORE-TNT1-Q1-Z1;  
CC IsoId=Q01295-1; Sequence=External;  
CC Name=2; Synonyms=BCORE-Q1-Z1;  
CC IsoId=Q01295-2; Sequence=External;  
CC Name=3; Synonyms=BCORE-Q2-Z1;  
CC IsoId=Q01295-3; Sequence=External;  
CC Name=4; Synonyms=BCORE-Z2;  
CC IsoId=Q01295-4; Sequence=External;  
CC Name=5; Synonyms=BCORE-NS-Z3;  
CC IsoId=Q01295-5; Sequence=External;  
CC -!- DEVELOPMENTAL STAGE: ACCUMULATES TO A HIGH LEVEL AT THE BEGINNING  
CC OF THE ECDYSONE RESPONSE, DURING THE METAMORPHOSIS OF IMAGINAL  
CC DISKS IN PUFF STAGE 1, AND ABRUPTLY DISAPPEARS AFTER SEVERAL  
CC HOURS.  
CC -!- INDUCTION: INDUCED AS A PRIMARY RESPONSE TO 20-HYDROXYECYDSONE IN  
CC THIRD INSTAR LARVAL IMAGINAL DISKS.  
CC -!- SIMILARITY: Contains 1 BTB/POZ domain.  
CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.  
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534  
CC TO 619 AND 656 TO 694 DUE TO FRAMESHIFTS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U51585; AB09760.1; ALT FRAME.  
CC EMBL; AL009146; CAAL5627.1; -  
CC EMBL; AB003421; AAF45647.1; -  
CC TRANSFAC; T01480; -  
CC FlyBase; FBgn0000210; br.  
CC InterPro; IPR000210; BTB\_POZ.  
CC InterPro; IPR007087; Znf\_C2H2.  
CC Pfam; PF00651; BTB; 1.  
CC Pfam; PF00096; zf-C2H2; 2.  
CC SMART; SM00225; BTB; 1.  
CC SMART; SM00353; Znf\_C2H2; 2.  
CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 2.  
CC PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 2.  
CC PROSITE; PS00097; BTB; 1.  
CC Nuclear protein; DNA-binding; Developmental protein;  
CC Zinc-finger; Metal-binding; Alternative splicing.  
FT DOMAIN 32 97  
FT ZN\_FING 710 733  
FT ZN\_FING 740 763  
FT DOMAIN 203 207  
FT DOMAIN 265 268  
FT DOMAIN 458 466  
FT DOMAIN 584 589  
FT DOMAIN 618 621  
FT DOMAIN 798 803

```

FT DOMAIN      821      833      POLY-ALA.
FT DOMAIN      862      867      POLY-GLN.
FT CONFLICT    436      436      G -> D (IN REF. 1).
FT CONFLICT    621      621      MISSING (IN REF. 1).
FT CONFLICT    624      624      A -> R (IN REF. 1).
FT CONFLICT    661      662      AV -> L (IN REF. 1).
FT CONFLICT    678      678      MISSING (IN REF. 1).
FT CONFLICT    722      723      KL -> NV (IN REF. 1).
SQ SEQUENCE    880 AA; 92305 MW; 500COAA38663AAF CRC64;

Query Match      79.3%; Score 23; DB 1; Length 880;
Best Local Similarity 37.5%; Pred. No. 7e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 HXXXHXXH 8
Db      625 HAHAAHAAH 632

RESULT 4
SYA_STRCO
ID SYA_STRCO STANDARD; PRT; 890 AA.
AC Q9KXP9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).
GN ALAS OR SC01501 OR SC905.25C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdono-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajadream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietorrek A., Woodward J., Bartell E.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC diphosphate + L-alanyl-tRNA(Ala).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL939109; CAB93381.1; -.
CC HAMAP; MF_00036; -.
CC InterPro; IPR003156; DHEA1.
CC InterPro; IPR002318; tRNA-synt_2c.
CC InterPro; IPR006193; tRNA_synt_Ala.
CC Pfam; PF02272; DHEA1; 1.
CC Pfam; PF01411; tRNA-synt_2c; 1.
CC PRINTS; PR00980; TRNASYNTHALA.
CC TIGRFAMs; TIGR00344; alaG; 1.
CC PROSITE; PS00860; AA tRNA LIGASE II ALA; 1.
KW Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding;
KW Complete proteome.

```

```

SQ SEQUENCE    890 AA; 95786 MW; 05B2FD563D35F4DF CRC64;

Query Match      79.3%; Score 23; DB 1; Length 890;
Best Local Similarity 37.5%; Pred. No. 7.1e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 HXXXHXXH 8
Db      573 HSAATHLTH 580

RESULT 5
GSRI_HUMAN
ID GSRI_HUMAN STANDARD; PRT; 1509 AA.
AC Q9NZM4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glioma tumor suppressor candidate region gene 1 protein.
GN GLUTSCR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RX MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
RA Portier B.P., Vekli K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
RA Scheithauer B.W., Louis D.N., Jenkins R.B.;
RA "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
RT region.";
RL Genomics 64:44-50(2000).
CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
CC placenta, skeletal muscle, and pancreas, and at lower levels in
CC lung, liver, and kidney.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF182077; AAF62874.1; -.
CC Genew; HGNC:4332; GLUTSCR1.
CC
CC MIM; 605690; -.
CC FT DOMAIN      37      45      POLY-GLY.
CC FT DOMAIN      884      889      POLY-PRO.
CC FT DOMAIN      1214     1225      POLY-SER.
CC FT DOMAIN      1282     1286      POLY-PRO.
CC FT DOMAIN      1294     1304      POLY-PRO.
SQ SEQUENCE    1509 AA; 152991 MW; 7C5144F443CE6821 CRC64;

Query Match      79.3%; Score 23; DB 1; Length 1509;
Best Local Similarity 37.5%; Pred. No. 1.1e+03;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 HXXXHXXH 8
Db      482 HSGAHSAAH 489

RESULT 6
YN78 YEAST
ID YN78 YEAST STANDARD; PRT; 52 AA.
AC P53820;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 6.0 kDa protein in COS1 5' region.
GN YNL338W OR N0170.

```

OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC NCB1\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Obermaier B., Piravandi E., Rinke M.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: TO YEAST YHR217C.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; 271614; CA96274.1; -;  
 DR EMBL; 271613; CA96273.1; -;  
 DR PIR; S63324; S63324.  
 DR SGD; S0005282; YNL338M.  
 KW Hypothetical protein.  
 SQ SEQUENCE 52 AA; 5951 MW; C1E4066D43E057A1 CRC64;  
 Query Match 75.9%; Score 22; DB 1; Length 52;  
 Best Local Similarity 37.5%; Pred. No. 90;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HXXHXXH 8  
 Db 38 HPHTHH 45  
 RESULT 7  
 ID Y034 TREPA STANDARD; PRT; 316 AA.  
 AC 083077;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Putative periplasmic metal-binding protein TP0034 precursor.  
 GN TP0034.  
 OS Treponema pallidum.  
 CC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
 CC NCB1\_TaxID=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Nichols;  
 RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Artlisch P., Bowman C., Corton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete."  
 RL Science 281:375-388 (1998).  
 CC -!- FUNCTION: PART OF AN ATP-DRIVEN TRANSPORT SYSTEM  
 CC TP0034/TP0035/TP0036 FOR A METAL. METAL-BINDING COMPONENT.  
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL SOLUTE-BINDING PROTEIN FAMILY  
 CC 9.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 DR EMBL; AE001188; AAC65029.1; -;  
 DR PIR; D71375; D71375.  
 DR TIGR; TP0034; -;  
 DR InterPro; IPR006128; Lipoprotein\_4.  
 DR InterPro; IPR006127; SBP\_bac\_9.  
 DR Pfam; PF01297; SBP\_bac\_9; 1.  
 DR PRINTS; PR00690; ADHESNFAMILY.  
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; UNKNOWN 1.  
 KW Hypothetical protein; Transport; Periplasmic; Metal-binding; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 316 PUTATIVE PERIPLASMIC METAL-BINDING  
 FT PROTEIN TP0034.  
 SQ SEQUENCE 316 AA; 35433 MW; 16051C219BC81AB CRC64;  
 Query Match 75.9%; Score 22; DB 1; Length 316;  
 Best Local Similarity 37.5%; Pred. No. 4.4e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HXXHXXH 8  
 Db 124 HTRGHTAH 131  
 RESULT 8  
 ADH1 BACST  
 ID ADH1 BACST STANDARD; PRT; 337 AA.  
 AC P12311;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Alcohol dehydrogenase (EC 1.1.1.1) (ADH-T).  
 GN ADHT.  
 OS Bacillus stearothermophilus.  
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
 CC NCB1\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
 RC STRAIN=NCA 1503;  
 RX MEDLINE=92138636; PubMed=1735726;  
 RA Sakoda H., Imanaka T.;  
 RT "Cloning and sequencing of the gene coding for alcohol dehydrogenase  
 RT of Bacillus stearothermophilus and rational shift of the optimum  
 RT pH."  
 RL J. Bacteriol. 174:1397-1402 (1992).  
 RN [2]  
 RP SEQUENCE OF 1-45.  
 RX MEDLINE=73229257; PubMed=4578954;  
 RA Bridgen J., Kolb E., Harris J.I.;  
 RT "Amino acid sequence homology in alcohol dehydrogenase."  
 RL FEBS Lett. 33:1-3 (1973).  
 RN [3]  
 RP SEQUENCE OF 34-54.  
 RX MEDLINE=79169263; PubMed=436831;  
 RA Jack R., Woenckhaus C., Harris J.I., Runswick M.J.;  
 RT "Identification of the amino acid residue modified in Bacillus  
 RT stearothermophilus alcohol dehydrogenase by the NAD+ analogue 4-(3-  
 RT bromoacetylpyridinio)butyldiphosphoadenosine."  
 RL Eur. J. Biochem. 93:57-64 (1979).  
 RN [4]  
 RP SEQUENCE OF 1-37; 188-197; 247-263 AND 324-336.  
 RC STRAIN=NCA 1503;  
 RX MEDLINE=94325354; PubMed=8049268;  
 RA Robinson G.A., Bailey C.J., Dowds B.C.A.;  
 RT "Gene structure and amino acid sequences of alcohol dehydrogenases of  
 RT Bacillus stearothermophilus."  
 RL Biochim. Biophys. Acta 1218:432-434 (1994).  
 CC -!- FUNCTION: THERMOSTABLE NAD(+)-DEPENDENT ALCOHOL DEHYDROGENASE.  
 CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
 CC NADH.  
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -!- ENZYME REGULATION: SUBSTRATE INHIBITION IS NOT OBSERVED WITH ANY

CC ALCOHOLS, AND THE ENZYME-NADH DISSOCIATION IS NOT CONSIDERED TO BE  
 CC A RATE-LIMITING STEP.  
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 CC family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D90421; BAA14411.1; -.  
 CC DR InterPro: IPR002328; ADH zinc.  
 CC DR Pfam; A42654; A42654.  
 CC DR InterPro; IPR002085; Adh\_zn family.  
 CC DR Pfam; PF00107; ADH\_ZINC\_N; 1.  
 CC DR PROSITE; PS00059; ADH\_ZINC; 1.  
 CC KW Oxidoreductase; Zinc; Metal-binding; NAD.  
 CC FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 CC FT METAL 61 61 ZINC 2 (CATALYTIC) (BY SIMILARITY).  
 CC FT METAL 92 92 ZINC 2 (BY SIMILARITY).  
 CC FT METAL 95 95 ZINC 2 (BY SIMILARITY).  
 CC FT METAL 98 98 ZINC 2 (BY SIMILARITY).  
 CC FT METAL 106 106 ZINC 2 (BY SIMILARITY).  
 CC FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 CC FT MUTAGEN 40 40 T-S: LITTLE DECREASE IN ACTIVITY.  
 CC FT MUTAGEN 43 43 H-R: HIGHER LEVEL OF ACTIVITY AT PH 9.  
 CC FT MUTAGEN 38 38 C-S: NO ACTIVITY.  
 CC FT MUTAGEN 40 40 T-S: NO ACTIVITY.  
 CC FT MUTAGEN 43 43 H-S: NO ACTIVITY.  
 CC FT CONFLICT 22 22 MISSING (IN REF. 2).  
 CC FT CONFLICT 33 33 MISSING (IN REF. 2).  
 CC FT CONFLICT 52 53 KP -> PK (IN REF. 3).  
 CC SQ SEQUENCE 337 AA; 36100 MW; B9B35A80E9B7A86 CRC64;  
 CC  
 CC Query Match 75.9%; Score 22; DB 1; Length 337;  
 CC Best Local Similarity 37.5%; Pred. No. 4.6e+02;  
 CC Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 CC  
 CC QY 1 HXXHXXH 8  
 CC DB 39 HTDLHAH 46  
 CC  
 CC RESULT 9  
 CC ADH2\_BACST STANDARD; PRT; 339 AA.  
 CC ID ADH2\_BACST STANDARD; PRT; 339 AA.  
 CC AC P42327;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 CC DE Alcohol dehydrogenase (EC 1.1.1.1) (ADH).  
 CC GN ADH.  
 CC OS Bacillus stearothermophilus.  
 CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
 CC OX NCBI\_TaxID=1422;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.  
 CC RC STRAIN=DSM 2334;  
 CC RX MEDLINE=94325354; PubMed=8049268;  
 CC RA Robinson G.A., Bailey C.J., Dows B.C.A.;  
 CC RT "Gene structure and amino acid sequences of alcohol dehydrogenases of  
 CC RL Bacillus stearothermophilus.";  
 CC RL Biochim. Biophys. Acta 1218:432-434(1994).  
 CC -!- FUNCTION: ACTIVE WITH PRIMARY ALCOHOLS, INCLUDING METHANOL.  
 CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
 CC NADH.  
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -!- ENZYME REGULATION: THE RATE-LIMITING STEP IS NADH RELEASE.  
 CC -!- CATABOLITE REPRESSION.  
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 CC family.

CC family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z25544; CAA80989.1; -.  
 CC DR InterPro: IPR002328; ADH zinc.  
 CC DR Pfam; PF00107; ADH\_ZINC\_N; 1.  
 CC DR PROSITE; PS00059; ADH\_ZINC; 1.  
 CC KW Oxidoreductase; Zinc; Metal-binding; NAD.  
 CC FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 CC FT METAL 61 61 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 CC FT METAL 92 92 ZINC 2 (BY SIMILARITY).  
 CC FT METAL 95 95 ZINC 2 (BY SIMILARITY).  
 CC FT METAL 98 98 ZINC 2 (BY SIMILARITY).  
 CC FT METAL 106 106 ZINC 2 (BY SIMILARITY).  
 CC FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 CC SQ SEQUENCE 339 AA; 36205 MW; 0EC33CE7287D7476 CRC64;  
 CC  
 CC Query Match 75.9%; Score 22; DB 1; Length 339;  
 CC Best Local Similarity 37.5%; Pred. No. 4.6e+02;  
 CC Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 CC  
 CC QY 1 HXXHXXH 8  
 CC DB 39 HTDLHAH 46  
 CC  
 CC RESULT 10  
 CC ADH3\_BACST STANDARD; PRT; 339 AA.  
 CC ID ADH3\_BACST STANDARD; PRT; 339 AA.  
 CC AC P42328;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 CC DE Alcohol dehydrogenase (EC 1.1.1.1) (ADH-HT).  
 CC OS Bacillus stearothermophilus.  
 CC OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
 CC OX NCBI\_TaxID=1422;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=NCIMB 12403 / LLD-R;  
 CC RX MEDLINE=94291628; PubMed=8020473;  
 CC RA Cannio R., Rossi M., Bartolucci S.;  
 CC RT "A few amino acid substitutions are responsible for the higher  
 CC RT thermostability of a novel NAD(+) dependent bacillar alcohol  
 CC RT dehydrogenase.";  
 CC RL Eur. J. Biochem. 222:345-352(1994).  
 CC -!- FUNCTION: THERMOSTABLE AND THERMOPHILIC NAD(+) DEPENDENT  
 CC ALCOHOL DEHYDROGENASE. BEARS MAINLY AN ETHANOL-DEHYDROGENASE  
 CC ACTIVITY.  
 CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
 CC NADH.  
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 CC family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z27089; CAA81612.1; -.



DR PIR; S45605; S45605.  
 DR HSSP; P28304; 100R.  
 DR InterPro; IPR002328; ADH\_zinc.  
 DR InterPro; IPR002085; Adh\_zn family.  
 DR Pfam; PF00107; ADH\_zinc\_N; 1.  
 DR PROSITE; PS00059; ADH\_ZINC; 1.  
 KW Oxidoreductase; Zinc; Metal-binding; NAD.  
 FT METAL 38 38 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 61 61 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 92 92 ZINC 2 (BY SIMILARITY).  
 FT METAL 95 95 ZINC 2 (BY SIMILARITY).  
 FT METAL 98 98 ZINC 2 (BY SIMILARITY).  
 FT METAL 106 106 ZINC 2 (BY SIMILARITY).  
 FT METAL 148 148 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 339 AA; 36338 MW; AED17E4A34163430 CRC64;  
 Query Match 75.9%; Score 22; DB 1; Length 339;  
 Best Local Similarity 37.5%; Pred. No. 4.6e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HXXXHXXH 8  
 Db 39 HTDJAHH 46  
 RESULT 11  
 ID\_FILIA\_HUMAN STANDARD; PRT; 416 AA.  
 AC P20930;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Filaggrin precursor (Fragment).  
 GN FLG.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89296901; PubMed=2740331;  
 RA McKinley-Grant L.J., Idler W.W., Bernstein I.A., Parry D.A.D.,  
 RA Cannizzaro L., Croce C.M., Huebner K., Lessner S.R., Steinert P.M.;  
 RT "Characterization of a cDNA clone encoding human filaggrin and  
 RT localization of the gene to chromosome region 1q21.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4848-4852(1989).  
 RN [2]  
 RP CITRULLINATION.  
 RX MEDLINE=9637438; PubMed=8780679;  
 RA Senshu T., Kan S., Ogawa H., Manabe M., Asaga H.;  
 RA "Preferential delamination of keratin K1 and filaggrin during the  
 RT terminal differentiation of human epidermis.";  
 RL Biochem. Biophys. Res. Commun. 225:712-719(1996).  
 CC -!- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES  
 CC DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING  
 CC TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.  
 CC -!- PFM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,  
 CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES  
 CC OF 324 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE  
 CC PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL  
 CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTOLYTICALLY CLEAVED.  
 CC -!- PFM: Undergoes delamination of some arginine residues  
 CC (citrullination).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M24355; AAA52454.1; -.

DR PIR; A32947; A32947.  
 DR Genew; HGNC:3748; FLG.  
 DR MIM; 135940; -.  
 DR GO; GO:0005882; C:intermediate filament; NAS.  
 DR GO; GO:0005198; F:structural molecule activity; NAS.  
 DR GO; GO:0007275; P:development; NAS.  
 DR InterPro; IPR003303; Filaggrin.  
 DR Pfam; PF03516; Filaggrin; 2.  
 DR PRINTS; PR00487; FILAGGRIN.  
 KW Phosphorylation; Citrullination; Developmental protein.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 416 AA; 44105 MW; DEEA3218BA043F32 CRC64;  
 Query Match 75.9%; Score 22; DB 1; Length 416;  
 Best Local Similarity 37.5%; Pred. No. 5.6e+02;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 HXXXHXXH 8  
 Db 99 HSGSHSH 106  
 RESULT 12  
 ID\_PFTB\_PEA STANDARD; PRT; 419 AA.  
 AC Q04903;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein farnesyltransferase beta subunit (EC 2.5.1.-) (CAAX  
 DE farnesyltransferase beta subunit) (RAS proteins prenyltransferase  
 DE beta) (Ftase-beta).  
 GN FTB.  
 OS Pisum sativum (Garden pea).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.  
 CX NCBI\_TaxID=3888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=cv. Alaska; TISSUE=Seedling;  
 RX MEDLINE=94105305; PubMed=8278509;  
 RA Yang Z., Cramer C.L., Watson J.C.;  
 RT "Protein farnesyltransferase in plants. Molecular cloning and  
 RT expression of a homolog of the beta subunit from the garden pea.";  
 RL Plant Physiol. 101:667-674(1993).  
 CC -!- FUNCTION: CATALYZES THE TRANSFER OF A FARNESYL MOIETY FROM  
 CC FARNESYL PYROPHOSPHATE TO A CYSTEINE AT THE FOURTH POSITION FROM  
 CC THE C-TERMINUS OF SEVERAL PROTEINS. THE BETA SUBUNIT IS  
 CC RESPONSIBLE FOR PEPTIDE-BINDING (BY SIMILARITY).  
 CC -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).  
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE PROTEIN PRENYLTRANSFERASE BETA SUBUNIT  
 CC FAMILY.  
 CC -!- SIMILARITY: Contains 5 PFTB repeats.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L08664; AAA33649.1; -.  
 DR PIR; JQ2254; JQ2254.  
 DR HSSP; Q02293; 1FT1.  
 DR InterPro; IPR001330; Prenyltrans.  
 DR Pfam; PF00432; prenyltrns; 5.  
 KW Transferase; Prenyltransferase; Repeat; Zinc.  
 FT REPEAT 68 109 PFTB 1.  
 FT REPEAT 119 160 PFTB 2.

```

FT REPEAT      167 208      PFTB 3.
FT REPEAT      215 256      PFTB 4.
FT REPEAT      329 371      PFTB 5.
FT METAL       241 241      ZINC (BY SIMILARITY).
FT METAL       243 243      ZINC (BY SIMILARITY).
FT METAL       359 359      ZINC (BY SIMILARITY).
SQ SEQUENCE    419 AA; 46793 MW; 4F040E094277D7C CRC64;

Query Match      75.9%; Score 22; DB 1; Length 419;
Best Local Similarity 37.5%; Pred. No. 5.6e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHHX 8
Db 291 HATSHIRH 298

RESULT 13
PROA NEIMB STANDARD; PRT; 420 AA.
ID PROA NEIMB
AC Q9JZG3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-
DE semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde
DE dehydrogenase) (GSA dehydrogenase).
GN PROA OR NM01068.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN SEQUENCE FROM N.A.
RP STRAIN=MC58 / Serogroup B;
RC MEDLINE=20175755; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC -!- FUNCTION: CATALYZES THE NADPH DEPENDENT REDUCTION OF L-GAMMA-
CC GLUTAMYL 5-PHOSPHATE INTO L-GLUTAMATE 5-SEMIALDEHYDE AND
CC PHOSPHATE. THE PRODUCT SPONTANEOUSLY UNDERGOES CYCLIZATION TO FORM
CC 1-PYRROLINE-5-CARBOXYLATE.
CC -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +
CC NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
CC -!- PATHWAY: Proline biosynthesis; second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002457; AAP62324.1; -.
CC TIGR; NMB1068; -.
CC HAMAP; MF_00412; -.
CC InterPro; IPR002086; Aldehyde dehydr.
CC TIGRFAMs; TIGR00407; proA; 1.
CC PROSITE; PS01223; proA; 1.
CC Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
CC SEQUENCE 420 AA; 45241 MW; A5D96CEDE50E87A2 CRC64;

```

---

```

DR PROSITE; PS01223; PROA; 1.
KW Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
SQ SEQUENCE 420 AA; 45256 MW; 00999B9CF6B1118 CRC64;

Query Match      75.9%; Score 22; DB 1; Length 420;
Best Local Similarity 37.5%; Pred. No. 5.6e+02;
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHHX 8
Db 335 HIETHSTH 342

RESULT 14
PROA STREP STANDARD; PRT; 420 AA.
ID PROA STREP
AC Q97R94;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-
DE semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde
DE dehydrogenase) (GSA dehydrogenase).
GN PROA OR SP0932.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC BAA-334 / TIGR4;
RC MEDLINE=21357209; PubMed=11463916;
RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radvane D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiolli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
CC -!- FUNCTION: Catalyzes the NADPH dependent reduction of L-gamma-
CC glutamyl 5-phosphate into L-glutamate 5-semialdehyde and
CC phosphate. The product spontaneously undergoes cyclization to form
CC 1-pyrroline-5-carboxylate.
CC -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +
CC NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
CC -!- PATHWAY: Proline biosynthesis; second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE007398; AAK75056.1; -.
CC TIGR; SP0932; -.
CC HAMAP; MF_00412; -.
CC InterPro; IPR002086; Aldehyde dehydr.
CC TIGRFAMs; TIGR00407; proA; 1.
CC PROSITE; PS01223; proA; 1.
CC Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
CC SEQUENCE 420 AA; 45241 MW; A5D96CEDE50E87A2 CRC64;

```

Query Match 75.9%; Score 22; DB 1; Length 420;  
Best Local Similarity 37.5%; Pred. No. 5.6e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
DB 335 HIESHSTH 342

## RESULT 15

FXH2 XENLA STANDARD; PRT; 427 AA.  
AC P32315;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE FXKH2 protein.  
GN FXKH2.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94074768; PubMed=8253274;  
RA Bolce M.E., Hemmati-Briuanlou A., Harland R.M.;  
RT "FXKH2, a Xenopus HNF-3 alpha homologue, exhibits both  
RT activin-inducible and autonomous phases of expression in early  
RT embryos.";  
RL Dev. Biol. 160:413-423(1993).  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- TISSUE SPECIFICITY: PRESENT IN THE VEGETAL POLE AND MARGINAL ZONE  
CC BUT NOT THE ANIMAL POLE OF GASTRULAE AND IN EQUAL LEVELS IN THE  
CC DORSAL AND VENTRAL HALVES OF BOTH GASTRULAE AND NEURULAE.  
CC -!- SIMILARITY: Contains 1 fork-head domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M93658; AAA17050.1; -;  
DR PIR; I51580; I51580.  
DR HSSP; Q63245; 2HFH.  
DR InterPro; IPR001766; TF\_Fork\_head.  
DR Pfam; PF00250; Fork\_head; 1.  
DR PRINTS; PR00053; FORKHEAD.  
DR ProDom; PD000425; TF\_Fork\_head; 1.  
DR SMART; SM00339; FH; 1.  
DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
KW DNA-binding; Nuclear protein.  
FT DNA\_BIND 156 247 FORK-HEAD.  
SQ SEQUENCE 427 AA; 46572 MW; 2D29A42AF960730C CRC64;

Query Match 75.9%; Score 22; DB 1; Length 427;  
Best Local Similarity 37.5%; Pred. No. 5.7e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
DB 333 HSLAHETH 340

Search completed: November 12, 2003, 09:35:39  
Job time : 11 secs

```

CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X62088; CAA43998.1; -
CC EMBL; X53180; CAA37321.1; -
CC PIR; JH0256; JH0256.
CC HSSP; P10845; 3BTA.
CC
CC MEROPS; M27.002; -
CC InterPro; IPR000395; Bontoxilysin.
CC InterPro; IPR006025; Zn_MTPeptide.
CC Pfam; PF01742; Peptidase M27; 1.
CC PRINTS; PR00760; BONTOXILYSIN.
CC ProDom; PD001963; Bontoxilysin; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
CC INIT MET 0
CC CHAIN 0
CC CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
CC CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
CC METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).
CC ACT_SITE 212 212 BY SIMILARITY.
CC METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).
CC DISULFID 411 425 INTERCHAIN (PROBABLE).
CC CONFLICT 229 229 K -> M (IN REF. 2).
CC
CC SEQUENCE 1250 AA; 143265 MW; 8171B5B2C312857 CRC64;
CC
CC Query Match 78.3%; Score 36; DB 1; Length 1250;
CC Best Local Similarity 75.0%; Pred. No. 37;
CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
CC QY 1 HDLIHVLH 8
CC Db 211 HELIHSIH 218
CC
CC RESULT 8
CC RECR_LISIN STANDARD; PRT; 198 AA.
CC AC Q927E9;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DE Recombination protein recR.
CC GN RECR OR LIN2850.
CC OS Listeria innocua.
CC OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
CC OX NCBI_TaxID=1642;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=CLIP 11262 / Serovar 6a;
CC RX MEDLINE=21537279; PubMed=11679669;
CC RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
CC Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
CC Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
CC Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
CC Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
CC Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
CC Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
CC Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
CC Nordisiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
CC Rammel B., Rose M., Schluter T., Simoes N., Tierrez A.,
CC Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
CC "Comparative genomics of Listeria species.";

```

```

RL Science 294:849-852(2001).
CC -!- FUNCTION: May play a role in DNA repair. It seems to be involved
CC in a recBC-independent recombinational process of DNA repair. It
CC may act with recF and recO (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL596173; CAC98076.1; -
CC PIR; AD1788; AD1788.
CC LiStiList; LIIN02850; -
CC HAMAP; MF_00017; -; 1.
CC InterPro; IPR003583; HHH 1.
CC InterPro; IPR000093; RecR.
CC InterPro; IPR006171; Toprim dom.
CC InterPro; IPR006154; Toprim_sub.
CC Pfam; PF02132; RecR; 1.
CC Pfam; PF01751; Toprim; 1.
CC SMART; SM00278; HHH1; 1.
CC SMART; SM00493; TOPRIM; 1.
CC TIGRFAMs; TIGR00615; recR; 1.
CC PROSITE; PS01300; RecR; 1.
CC DNA repair; DNA recombination; Zinc-finger; Complete proteome.
CC ZN_FING 57 72 C4-TYPE (POTENTIAL).
CC SEQUENCE 198 AA; 21996 MW; 8A82E1A16415DDEF CRC64;
CC
CC Query Match 73.9%; Score 34; DB 1; Length 198;
CC Best Local Similarity 75.0%; Pred. No. 12;
CC Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CC
CC QY 1 HDLIHVLH 8
CC Db 101 HGLYHVLH 108
CC
CC RESULT 9
CC RECR_LISMO STANDARD; PRT; 198 AA.
CC AC Q8Y3X7;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Recombination protein recR.
CC GN RECR OR IMO2702.
CC OS Listeria monocytogenes.
CC OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
CC OX NCBI_TaxID=1639;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=BGD-e / Serovar 1/2a;
CC RX MEDLINE=21537279; PubMed=11679669;
CC RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
CC Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
CC Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
CC Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
CC Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
CC Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
CC Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
CC Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
CC Nordisiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
CC Rammel B., Rose M., Schluter T., Simoes N., Tierrez A.,
CC Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
CC "Comparative genomics of Listeria species.";
CC Science 294:849-852(2001).
CC -!- FUNCTION: May play a role in DNA repair. It seems to be involved
CC in a recBC-independent recombinational process of DNA repair. It
CC may act with recF and recO (By similarity).

```

```

CC CC -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; AL591984; CAD00915.1; -
CC CC PIR; AE1412; AE1412.
CC CC List; LMO02702; -
CC CC HAMAP; MF_00017; -; 1.
CC CC InterPro; IPR0003583; HHH 1.
CC CC InterPro; IPR000093; RecR.
CC CC InterPro; IPR006171; Toprim_dom.
CC CC InterPro; IPR006154; Toprim_sub.
CC CC Pfam; PF02132; RecR; 1.
CC CC Pfam; PF01751; Toprim; 1.
CC CC SMART; SM00278; HhH1; 1.
CC CC SMART; SM00493; TOPRIM; 1.
CC CC TIGRFAMs; TIGR00615; recR; 1.
CC CC DNA repair; DNA recombination; Zinc-finger; Complete proteome.
KW DNA repair; C4-TYPE (POTENTIAL).
FT ZN FING 57 72 C4-TYPE (POTENTIAL).
SQ SEQUENCE 198 AA; 21934 MW; E542E27BC3D05036 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 198;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
DB 101 HGLYHVLH 108

RESULT 10
RECR STRN STANDARD; PRT; 198 AA.
AC Q9ZHC4;
AT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Recombination protein recR.
GN RECR OR RECM OR SPI672 OR SPI1616.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN [1]_TaxID=1313, 171101;
RP SEQUENCE FROM N.A.
RC STRAIN=G54 / Type 19F;
RX MEDLINE=99061199; PubMed=9846742;
RA Massidda O., Anderluzzi D., Friedli L., Feger G.;
RT "Unconventional organization of the division and cell wall gene
RT cluster of Streptococcus pneumoniae.";
RL Microbiology 144:3069-3078(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=339, and PN94-661;
RX MEDLINE=20073037; PubMed=10605111;
RA Erhardt M.C., Spratt B.G.;
RT "Extensive variation in the ddl gene of penicillin-resistant
RT Streptococcus pneumoniae results from a hitchhiking effect driven by
RT the penicillin-binding protein 2b gene.";
RL Mol. Biol. Evol. 16:1687-1695(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,

```

```

RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geisinger C.,
RA Gilmour R.J., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nickas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostack P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
CC -!- FUNCTION: MAY PLAY A ROLE IN DNA REPAIR. IT SEEMS TO BE INVOLVED
CC -!- IN AN RECOMBINATION RECOMBINATIONAL PROCESS OF DNA REPAIR. IT
CC MAY ACT WITH RECJ AND RECQ (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; AF068901; AAC95434.1; -
CC CC EMBL; AJ243056; CAB64474.1; -
CC CC EMBL; AJ243057; CAB64478.1; -
CC CC EMBL; AE007460; AAK75751.1; -
CC CC EMBL; AE008520; AAL00320.1; -
CC CC PIR; C98061; C98061.
CC CC PIR; F95194; F95194.
CC CC TIGR; SP1672; -; 1.
CC CC HAMAP; MF_00017; -; 1.
CC CC InterPro; IPR000093; RecR.
CC CC InterPro; IPR006171; Toprim_dom.
CC CC InterPro; IPR006154; Toprim_sub.
CC CC Pfam; PF02132; RecR; 1.
CC CC Pfam; PF01751; Toprim; 1.
CC CC SMART; SM00493; TOPRIM; 1.
CC CC TIGRFAMs; TIGR00615; recR; 1.
CC CC PROSITE; PS01300; RECQ; 1.
KW DNA repair; DNA recombination; Zinc-finger; Complete proteome.
FT ZN FING 57 72 C4-TYPE (POTENTIAL).
SQ SEQUENCE 198 AA; 21689 MW; FC6F0B98E393752 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 198;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
DB 101 HGLYHVLH 108

RESULT 11
SODC HAEDU STANDARD; PRT; 199 AA.
ID SODC HAEDU STANDARD;
AC Q59452; Q59449; Q59453;
DT 15-DEC-1998 (Rel. 37, Created)

```



```

Query Match      73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVL 7
DB      155 HDLIHVL 161

RESULT 13
EX7L_SALTY STANDARD; PRT; 449 AA.
AC Q8Z4Q1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR STY2753 OR T0345.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Murgall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyonianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 195:2330-2337(2003).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL627275; CAD02714.1; -;
CC EMBL; AE016835; AAC06065.1; -;
CC HAMAP; MF_00378; -; 1.
CC InterPro; IPR003753; Exonuc_VII_L.
CC InterPro; IPR004365; tRNA_anti-.
CC Pfam; PF02601; Exonuc_VII_L; 1.
CC Pfam; PF01336; tRNA_anti-; 1.
CC TIGRFAMs; TIGR00237; xsea; 1.
CC KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
CC SEQUENCE 449 AA; 50720 MW; 511957DEC878F5D2 CRC64;

Query Match      73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVL 7
DB      154 HDLIHVL 160

RESULT 14
EX7L_SALTY STANDARD; PRT; 449 AA.
AC Q8ZN58;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR STM2512.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008813; AAL21406.1; -;
CC HAMAP; MF_00378; -; 1.
CC InterPro; IPR003753; Exonuc_VII_L.
CC InterPro; IPR004365; tRNA_anti-.
CC Pfam; PF02601; Exonuc_VII_L; 1.
CC Pfam; PF01336; tRNA_anti-; 1.
CC TIGRFAMs; TIGR00237; xsea; 1.
CC KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
CC SEQUENCE 449 AA; 50613 MW; 85356CE8560E161E CRC64;

Query Match      73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVL 7
DB      154 HDLIHVL 160

```

```
RESULT 15
EX7L ECO57 STANDARD; PRT; 456 AA.
AC Q8XAB0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR Z3773 OR EC93371.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11208551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimallanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB005480; AAG57620.1; -
DR EMBL; AF002561; BAB36794.1; -
DR FIR; C91050; C91050.
DR HAMAP; MF_00378; -; 1.
DR InterPro; IPR003753; Exonuc VII L.
DR InterPro; IPR004365; tRNA_anti_.
DR Pfam; PF02601; Exonuc VII_L; 1.
DR Pfam; PF01336; tRNA_anti; 1.
DR TIGRFAMs; TIGR00237; xsea; 1.
KW Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 456 AA; 51734 MW; 174EAE7F72EB3C37 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 456;
Best Local Similarity 71.4%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLHVL 7
|::|||
```

Db 154 HDLHVL 160

Search completed: November 12, 2003, 09:35:40  
Job time : 11 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:32:27 ; Search time 28 seconds  
(without alignments)  
73.729 Million cell updates/sec

Title: US-10-064-903-2

Perfect score: 46

Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	91.3	451	2 Q9R631	Q9r631 clostridium
2	42	91.3	1268	2 Q45851	Q45851 clostridium
3	42	91.3	1291	2 Q9ZAJ3	Q9ZAJ3 clostridium
4	42	91.3	1291	2 Q93G71	Q93G71 clostridium
5	42	91.3	1291	2 Q933K0	Q933K0 clostridium
6	42	91.3	1291	2 Q88077	Q88077 clostridium
7	42	91.3	1291	2 Q8GR96	Q8GR96 clostridium
8	42	91.3	1310	2 Q93N27	Q93N27 clostridium
9	39	84.8	397	16 Q92ML7	Q92ML7 rhizobium m
10	38	82.6	925	16 Q9JS16	Q9JS16 chlamydia p
11	38	82.6	925	16 Q9Z8P5	Q9Z8P5 chlamydia p
12	38	82.6	1278	2 Q57236	Q57236 clostridium
13	38	82.6	1280	2 Q9ZAJ5	Q9ZAJ5 clostridium
14	37	80.4	349	5 Q8IJV0	Q8IJV0 plasmodium
15	37	80.4	707	3 Q8X008	Q8X008 neurospora
16	36	78.3	105	10 Q8VXL7	Q8VXL7 fagus sylv

17	36	78.3	232	16	Q8X829	Q8x829 escherichia
18	36	78.3	237	13	Q8AWC9	Q8awc9 cyprinus ca
19	36	78.3	241	10	Q8VXL6	Q8vxl6 fagus sylv
20	36	78.3	489	9	Q9AZH2	Q9azh2 bacterioph
21	36	78.3	489	16	Q9CEA2	Q9cea2 lactococcus
22	36	78.3	773	10	Q9SHP2	Q9shp2 arabidopsis
23	36	78.3	1251	2	Q9K395	Q9k395 clostridium
24	36	78.3	1252	2	Q8K2M3	Q8k2m3 clostridium
25	36	78.3	1255	2	Q9FAR6	Q9far6 clostridium
26	35	76.1	312	16	Q26068	Q26068 helicobacte
27	35	76.1	312	16	Q9ZJ59	Q9zj59 helicobacte
28	35	76.1	426	5	Q8MPY4	Q8mypy4 caenorhabdi
29	35	76.1	431	5	Q9XTZ9	Q9xtz9 caenorhabdi
30	35	76.1	500	16	Q9RVQ8	Q9rvq8 deiococcus
31	35	76.1	679	16	Q97SL8	Q97sl8 streptococc
32	35	76.1	737	16	Q8DR83	Q8dr83 streptococc
33	34	73.9	129	3	Q8TFI7	Q8tfi7 ustilago vi
34	34	73.9	198	2	Q9RCP8	Q9rcp8 streptococc
35	34	73.9	198	2	Q9RCQ5	Q9rcq5 streptococc
36	34	73.9	198	2	Q9RCR0	Q9rcr0 streptococc
37	34	73.9	198	2	Q9R2M1	Q9r2m1 streptococc
38	34	73.9	198	2	Q9RCQ2	Q9rcq2 streptococc
39	34	73.9	199	16	Q8DV99	Q8dv99 streptococc
40	34	73.9	222	16	Q8EE87	Q8ee87 shewanella
41	34	73.9	225	5	Q811V8	Q811v8 plasmodium
42	34	73.9	261	5	Q81FO1	Q81fq1 plasmodium
43	34	73.9	299	17	Q96YN4	Q96yn4 sulfobolus
44	34	73.9	385	10	Q8H3N0	Q8h3n0 oryza sativ
45	34	73.9	458	16	Q8FF64	Q8ff64 escherichia

#### ALIGNMENTS

#### RESULT 1

Q9R631 ID Q9R631 PRELIMINARY; PRT; 451 AA.  
AC Q9R631;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Neurotoxin type B light chain, BONT/B.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92340509; PubMed=1634516;  
RA Kurazono H., Mochida S., Binz T., Eisel U., Quanz M., Grebenstein O.,  
RA Wernars K., Poulain B., Tauc L., Niemann H.;  
RT "Minimal essential domains specifying toxicity of the light chains of  
RT tetanus toxin and botulinum neurotoxin type A.";  
RL J. Biol. Chem. 267:14721-14729 (1992).  
DR HSSP; P10845; 3BETA.  
DR InterPro; IPR000395; Bontoxilysin.  
DR InterPro; IPR006025; Zn\_Mtpeptidse.  
DR Pfam; PF01742; Peptidase\_M27; 1.  
DR PRINTS; PR00760; BONTOXILYSIN.  
DR PRODOM; PD001963; Bontoxilysin; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
SQ SEQUENCE 451 AA; 51943 MW; 6C79FD488653EA71 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 451;  
Best Local Similarity 87.5%; Pred. No. 6.3;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8

Db 230 HDLIHVLH 237

#### RESULT 2

```
Q45851
ID Q45851 PRELIMINARY; PRT; 1268 AA.
AC Q45851;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type F.
GN BONT /F.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1561;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93252228; PubMed=8486245;
RA Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA Richardson P.T.;
RT "Nucleotide sequence of the gene coding for Clostridium botulinum type F
RT neurotoxin: Comparison with other clostridial neurotoxins.";
RL FEMS Microbiol. Lett. 108:175-182(1993).
DR EMBL; X68262; CAA48329.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTPeptide.
DR Pfam; PF01742; Peptidase_M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1268;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 219 HELIHLVH 226

RESULT 3
Q9ZAJ8 PRELIMINARY; PRT; 1291 AA.
AC Q9ZAJ8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Bont protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 3281;
RX MEDLINE=9840323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13630; CRA73968.1; -.
DR HSSP; P10845; 3BTA.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTPeptide.
DR Pfam; PF01742; Peptidase_M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150840 MW; E4D3B0E46AB2E735 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 HDLIHVLH 8
Db 230 HELIHLVH 237

RESULT 4
Q93G71 PRELIMINARY; PRT; 1291 AA.
AC Q93G71;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type B.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1436;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF295926; AAK97132.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTPeptide.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150824 MW; D7CA07BAE2EB8CD2 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 230 HELIHLVH 237

RESULT 5
Q933KO PRELIMINARY; PRT; 1291 AA.
AC Q933KO;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type B cryptic neurotoxin.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=593; and 588;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF300466; AAL11499.1; -.
DR EMBL; AF300465; AAL11498.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTPeptide.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150843 MW; 7AC1737B0FASA151 CRC64;
```

Query Match 91.3%; Score 42; DB 2; Length 1291;  
 Best Local Similarity 87.5%; Pred. No. 18;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 |:|||||  
 Db 230 HELIHLVH 237

## RESULT 6

Q08077 PRELIMINARY; PRT; 1291 AA.  
 ID Q08077  
 AC Q08077  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE BONT/B.  
 GN BONT/B.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Eklund 17B ATCC25765;  
 RX MEDLINE=94122659; PubMed=7764370;  
 RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;  
 RT "Nucleotide sequence of the gene coding for non-proteolytic  
 Clostridium botulinum type B neurotoxin: comparison with other  
 Clostridial neurotoxins";  
 RL Curr. Microbiol. 28:101-110(1994).  
 DR EMBL; X71343; CAA50482.1; -.  
 DR HSSP; P10845; 3BTA.  
 DR MEROPS; M27.002; -.  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR006025; Zn\_Mtpeptidse.  
 DR Pfam; PF01742; Peptidase M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 SQ SEQUENCE 1291 AA; 150513 MW; 71BCAFE23D69FAAA CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;  
 Best Local Similarity 87.5%; Pred. No. 18;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 |:|||||  
 Db 230 HELIHLVH 237

## RESULT 7

Q08096 PRELIMINARY; PRT; 1291 AA.  
 ID Q08096  
 AC Q08096  
 DT 01-MAR-2003 (TREMELrel. 23, Created)  
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE Neurotoxin.  
 GN BONTB.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Eklund 17B ATCC25765;  
 RX MEDLINE=94122659; PubMed=7764370;  
 RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;  
 RT "Nucleotide sequence of the gene coding for non-proteolytic  
 Clostridium botulinum type B neurotoxin: comparison with other  
 Clostridial neurotoxins";  
 RL Curr. Microbiol. 28:101-110(1994).  
 DR EMBL; X71343; CAA50482.1; -.  
 DR HSSP; P10845; 3BTA.  
 DR MEROPS; M27.002; -.  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR006025; Zn\_Mtpeptidse.  
 DR Pfam; PF01742; Peptidase M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 SQ SEQUENCE 1291 AA; 150513 MW; 71BCAFE23D69FAAA CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;  
 Best Local Similarity 87.5%; Pred. No. 18;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 |:|||||  
 Db 230 HELIHLVH 237

## RESULT 8

Q09327 PRELIMINARY; PRT; 1310 AA.  
 ID Q09327  
 AC Q09327  
 DT 01-DEC-2001 (TREMELrel. 19, Created)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE Tetanus toxin (Fragment).  
 OS Clostridium tetani.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Shumin Z., Dianliang L.;  
 RT "Cloning and sequence analysis of tetanus toxin gene";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF389424; AAK72984.2; -.  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR006025; Zn\_Mtpeptidse.  
 DR Pfam; PF01742; Peptidase M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 FT NON TER 1  
 FT NON TER 1310 1310  
 SQ SEQUENCE 1310 AA; 150316 MW; 9EADDC914418E450 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1310;  
 Best Local Similarity 87.5%; Pred. No. 18;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 |:|||||  
 Db 234 HELIHLVH 241

## RESULT 9

Q092ML7 PRELIMINARY; PRT; 397 AA.  
 ID Q092ML7  
 AC Q092ML7  
 DT 01-DEC-2001 (TREMELrel. 19, Created)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE Putative deaminase OR deamidase protein.  
 GN R03596 OR SMC02420.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,  
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramperger U.,  
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 Sinorhizobium meliloti strain 1021.";

```

RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
DR EMBL; AL591791; CAC47175.1; -.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR001130; TatD_DNase.
DR Pfam; PF01979; Amidohydro_1; 1.
DR PROSITE; PS01137; TATD_1; 1.
KW Complete proteome.
SQ SEQUENCE 397 AA; 43054 MW; 87D5F69C499CBE02 CRC64;
Query Match 84.8%; Score 39; DB 16; Length 397;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
DB 326 HDLEHVLH 333
RESULT 10
Q9J516 PRELIMINARY; PRT; 925 AA.
AC Q9J516;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE CT234 hypothetical protein.
GN CFU0293 OR CP0465.
OS Chlamydia pneumoniae [Chlamydia pneumoniae].
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolchay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314 (2000).
DR EMBL; AF002208; AAF38302.1; -.
DR EMBL; AP002546; BAA98503.1; -.
DR TIGR; CP0465; -.
SQ SEQUENCE 925 AA; 105601 MW; 61E8941E7C8FD620 CRC64;
Query Match 82.6%; Score 38; DB 16; Length 925;
Best Local Similarity 62.5%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
DB 538 HDLHITH 545
RESULT 11
Q9Z8P5 PRELIMINARY; PRT; 925 AA.
AC Q9Z8P5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CT234 hypothetical protein.
GN CPN0293.
OS Chlamydia pneumoniae [Chlamydia pneumoniae].
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389 (1999).
DR EMBL; AE001614; AAD18442.1; -.
KW Complete proteome.
SQ SEQUENCE 925 AA; 105615 MW; 98E6098E7C8FD37D CRC64;
Query Match 82.6%; Score 38; DB 16; Length 925;
Best Local Similarity 62.5%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
DB 538 HDLHITH 545
RESULT 12
Q57236 PRELIMINARY; PRT; 1278 AA.
ID Q57236
AC Q57236; Q45863;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Botulinum neurotoxin type F (BONT/F protein).
GN BONT/F.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 10281;
RA Hutson R.A., Collins M.D.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 635-1000 FROM N.A.
RC STRAIN=NCTC 1028;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F."
RL J. Clin. Microbiol. 31:2255-2262 (1993).
RN [4]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN=LANGELEND;
RX MEDLINE=98404102; PubMed=9732534;
RA East A.K., Bhandari M., Hielm S., Collins M.D.;
RT "Analysis of the botulinum neurotoxin type F gene clusters in
RT proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT baratii."
RL Curr. Microbiol. 37:262-268 (1998).
DR EMBL; X81714; CAA57358.1; -.
DR EMBL; L35496; AAA23210.1; -.
DR EMBL; X70821; CAA50152.1; -.
DR EMBL; X93064; CAA67512.1; -.
DR HSSP; P10845; BETA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_Mtpeptidase.
DR Pfam; PF01742; Peptidase_M27; 1.
```

```

DR PRINTS; PR00760; BONTOKILYSIN.
DR PRODOM; PD001963; BONTOKILYSIN; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin.
SQ SEQUENCE 1278 AA; 147073 MW; 14B1318431D6918 CRC64;

Query Match      82.6%; Score 38; DB 2; Length 1278;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 227 HELIHALH 234

RESULT 13
Q9ZAJ5 PRELIMINARY; PRT; 1280 AA.
AC Q9ZAJ5;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Bont protein.
DE BONT.
GN Bont.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281;
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RA "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins".
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13631; CAA73972.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontokilysin.
DR InterPro; IPR006025; Zn_Mpeptidase.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOKILYSIN.
DR PRODOM; PD001963; BONTOKILYSIN; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1280 AA; 147487 MW; D0F748976BEC222C CRC64;

Query Match      82.6%; Score 38; DB 2; Length 1280;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 227 HELIHALH 234

RESULT 14
Q81JVO PRELIMINARY; PRT; 349 AA.
AC Q81JVO;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF10_0092.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;

RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shalimov S.J., Suh B., Peterson J., Anguoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.,
RA "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014830; AAN35290.1; -.
KW Hypothetical protein.
SQ SEQUENCE 349 AA; 40775 MW; 9D7C20FEACEB4464 CRC64;

Query Match      80.4%; Score 37; DB 5; Length 349;
Best Local Similarity 50.0%; Pred. No. 43;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 96 HELVHIVH 103

RESULT 15
Q8X008 PRELIMINARY; PRT; 707 AA.
AC Q8X008;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Related to hydroxyproline-rich glycoprotein.
GN B2H20.050.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hobeisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Manthaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL669988; CAD21077.1; -.
DR InterPro; IPR002965; P rich_extensn.
DR PRINTS; PR01217; PRICHEXTENSN.
SQ SEQUENCE 707 AA; 77817 MW; C49BC3C1A18D83F5 CRC64;

Query Match      80.4%; Score 37; DB 3; Length 707;
Best Local Similarity 62.5%; Pred. No. 87;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 158 HDLIHVLH 165

Search completed: November 12, 2003, 09:36:51
Job time : 31 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:32:27 ; Search time 28 Seconds  
(without alignments)  
73.729 Million cell updates/sec

Title: US-10-064-903-1

Perfect score: 29

Sequence: 1 HXXHXXH 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP\_TREMBL\_23.\*  
2: sp\_archaea.\*  
3: sp\_bacteria.\*  
4: sp\_fungi.\*  
5: sp\_human.\*  
6: sp\_invertebrate.\*  
7: sp\_mammal.\*  
8: sp\_mhc.\*  
9: sp\_organelle.\*  
10: sp\_phage.\*  
11: sp\_plant.\*  
12: sp\_rodent.\*  
13: sp\_virus.\*  
14: sp\_vertebrate.\*  
15: sp\_unclassified.\*  
16: sp\_virus.\*  
17: sp\_bacteriophage.\*  
18: sp\_bacteriophage.\*  
19: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	79.3	119	16	Q92K02
2	23	79.3	152	17	Q9Y47
3	23	79.3	177	5	Q9XWQ9
4	23	79.3	204	2	Q9XW65
5	23	79.3	240	16	Q9PFU7
6	23	79.3	324	2	Q9F502
7	23	79.3	327	2	Q9FD42
8	23	79.3	327	16	Q8YU45
9	23	79.3	341	5	Q8YU45
10	23	79.3	341	5	Q8YU45
11	23	79.3	342	5	Q18888
12	23	79.3	382	16	Q54173
13	23	79.3	402	2	Q9KX27
14	23	79.3	412	13	Q98U15
15	23	79.3	466	10	Q947K4
16	23	79.3	472	5	Q09657

17	79.3	534	5	Q8MTV4	Q8mtv4 plasmodium
18	79.3	536	5	Q8I3N3	Q8i3n3 plasmodium
19	79.3	596	16	Q9L1Q1	Q9l1q1 streptomyces
20	79.3	606	16	Q31566	Q31566 bacillus su
21	79.3	610	2	Q52961	Q52961 bacillus su
22	79.3	635	4	Q13476	Q13476 homo sapien
23	79.3	676	5	Q95WV0	Q95wv0 drosophila
24	79.3	677	6	Q28256	Q28256 canis famil
25	79.3	690	5	Q8IM21	Q8im21 plasmodium
26	79.3	696	5	Q9VDP3	Q9vdp3 drosophila
27	79.3	713	16	Q8XR50	Q8xr50 ralstonia s
28	79.3	826	10	Q9SCU4	Q9scu4 arabidopsis
29	79.3	826	10	Q9SCU3	Q9scu3 arabidopsis
30	79.3	899	3	Q8NIZ0	Q8niz0 neurospora
31	79.3	1031	5	Q8MT64	Q8mt64 drosophila
32	79.3	1059	5	Q9VZ52	Q9vz52 drosophila
33	79.3	1182	5	Q9VXL1	Q9vxl1 drosophila
34	79.3	1226	5	Q9V4U3	Q9v4u3 drosophila
35	79.3	1359	5	Q9VX26	Q9vx26 drosophila
36	79.3	3036	4	Q8TDJ6	Q8tdj6 homo sapien
37	79.3	3469	5	Q9U4I2	Q9u4i2 drosophila
38	79.3	3604	5	Q9VYK0	Q9vyk0 drosophila
39	79.3	4360	3	Q9UVN5	Q9uvn5 alternaria
40	75.9	56	3	Q8TGJ7	Q8tgj7 saccharomyc
41	75.9	57	16	Q8PE57	Q8pe57 xanthomonas
42	75.9	61	16	Q8ZE32	Q8ze32 yersinia pe
43	75.9	77	12	Q8QN71	Q8qn71 ectocarpus
44	75.9	93	7	Q9GJ30	Q9gj30 homo sapien
45	75.9	93	7	Q9GJ32	Q9gj32 homo sapien

#### ALIGNMENTS

RESULT 1

Q92K02 PRELIMINARY; PRT; 119 AA.  
AC Q92K02;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Hypothetical transmembrane protein SMC01986.  
GN R02520 OR SMC01986.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430;  
RA Capella D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,  
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramperger U.,  
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;  
RT "Analysis of the chromosome sequence of the legume symbiont  
Sinorhizobium meliloti strain 1021.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
DR EMBL; AL591791; CAC47099.1;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 119 AA; 13504 MW; CFPAL042CA34D6A9 CRC64;

Query Match 79.3%; Score 23; DB 16; Length 119;  
Best Local Similarity 37.5%; Pred. No. 8e+02;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXHXXH 8

DB 11 HAAHTEH 18

RESULT 2

Db 140 HTTVHSH 147

RESULT 4

ID Q9WX65 PRELIMINARY; PRT; 204 AA.

AC Q9WX65;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE BGL protein (Fragment).

GN BGL.

GN Acetobacter xylinus.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;

OC Acetobacteraceae; Gluconacetobacter.

OX NCBI\_TaxID=28448;

RN [1]

RC STRAIN=JCM7664;

RC Umeda Y., Hirano A., Hon-nami K., Kunito S., Akiyama H., Onizuka T.,

RA Ikeuchi M., Inoue Y.;

RT "Conversion of CO2 into cellulose by gene manipulation of microalgae:

RT cloning of cellulose synthase genes from Acetobacter xylinum.";

RL (In) Inui T., Anpo M., Izui K., Yanagida S., Yamaguchi T. (eds.);

RL Advances in chemical conversions for mitigating carbon dioxide,

RL pp.114:653-656, Elsevier Science, Amsterdam (1998).

RL EMBL; AB015802; BAA77589.1; -

DR InterPro; IPR001764; Glyco\_hydro\_3N.

DR Pfam; PF00933; Glyco\_hydro\_3; 1.

DR PRINTS; PR00133; GLYDRLASE3.

FT NON\_TER 204

SQ SEQUENCE 204 AA; 21146 MW; 2CD1050D8E2E720F CRC64;

Query Match 79.3%; Score 23; DB 2; Length 204;

Best Local Similarity 37.5%; Pred. No. 1.2e+03;

Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps

QY 1 HXXHXXHXH 8

DB 21 HDAAHNAH 28

RESULT 5

Q9PFU7 PRELIMINARY; PRT; 240 AA.

ID Q9PFU7

AC Q9PFU7;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE GMP synthase.

DE XF0560.

OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xylella.

OX NCBI\_TaxID=2371;

RN [1]

RC SEQUENCE FROM N.A.

RC STRAIN=9a5C;

RC MEDLINE=20365717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,

RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,

RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,

RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,

RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.N., Costa-Neto C.H.,

RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,

RA Facinanci A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,

RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,

RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,

RA Ho P.L., Honeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,

RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,

RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

RA Machado A.M., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,

RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tuhako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Seubal J.C.;  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*";  
 RL Nature 406:151-159(2000).  
 DR EMBL; AE003903; AAF83370.1; -.  
 DR InterPro; IPR000991; GATase\_1.  
 DR Pfam; PF00117; GATase; 1.  
 DR PRINTS; PR00096; GATASE.  
 DR PROSITE; PS00442; GATASE\_TYPE\_I; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 240 AA; 26350 MW; FF81B5EE1EBE35 CRC64;

Query Match 79.3%; Score 23; DB 16; Length 240;  
 Best Local Similarity 37.5%; Pred. No. 1.4e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8

Db 142 HFSAHATH 149

## RESULT 6

Q9F502 Q9F502 PRELIMINARY; PRT; 324 AA.  
 AC Q9F502;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Cytochrome c oxidase subunit 2 (EC 1.9.3.1).  
 GN COXB2.  
 OS Anabaena variabilis.  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
 OX NCBI\_TaxID=1172;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=ATCC29413;  
 RA Pills D., Schmetterer G.;  
 RT "A second cytochrome c oxidase from the cyanobacterium *Anabaena* sp.  
 strain ATCC29413, up-regulated under nitrogen fixing conditions.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 CC EMBL; AJ296086; CAC12660.1; -.  
 DR InterPro; IPR001505; Copper\_CUA.  
 DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
 DR Pfam; PF00116; COX2; 1.  
 DR Pfam; PF02790; COX2.TM; 1.  
 DR PRINTS; PR01166; CYCOXIDASEII.  
 DR PRODOM; PD000131; Copper\_CUA; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 KW Copper; Oxidoreductase; Transmembrane.  
 SQ SEQUENCE 324 AA; 34739 MW; E16B6CC160899F72 CRC64;

Query Match 79.3%; Score 23; DB 2; Length 324;  
 Best Local Similarity 37.5%; Pred. No. 1.8e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8

Db 126 HASAHVAH 133

RESULT 7  
 Q9FD42 Q9FD42 PRELIMINARY; PRT; 327 AA.  
 AC Q9FD42;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Cytochrome c oxidase subunit II.  
 GN CTA2.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=7120;  
 RA Jones K.M., Buikema W.J., Haselkorn R.;  
 RT "Characterization of a heterocyst-specific cytochrome c oxidase operon  
 in *Anabaena* PCC7120.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 CC EMBL; AF291994; AAG01550.1; -.  
 DR InterPro; IPR001505; Copper\_CUA.  
 DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
 DR Pfam; PF00116; COX2; 1.  
 DR Pfam; PF02790; COX2.TM; 1.  
 DR PRINTS; PR01166; CYCOXIDASEII.  
 DR PRODOM; PD000131; Copper\_CUA; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 KW Copper; Oxidoreductase; Transmembrane.  
 SQ SEQUENCE 327 AA; 35003 MW; 92324730EB7A92F8 CRC64;

Query Match 79.3%; Score 23; DB 2; Length 327;  
 Best Local Similarity 37.5%; Pred. No. 1.8e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 HXXXHXXH 8

Db 126 HASAHVAH 133

## RESULT 8

Q8YU45 Q8YU45 PRELIMINARY; PRT; 327 AA.  
 AC Q8YU45;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Cytochrome c oxidase subunit II.  
 GN COXB OR ALR2514.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shampo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium *Anabaena* sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AP003589; BAB74213.1; -.  
 DR InterPro; IPR001505; Copper\_CUA.  
 DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
 DR Pfam; PF00116; COX2; 1.  
 DR Pfam; PF02790; COX2.TM; 1.  
 DR PRINTS; PR01166; CYCOXIDASEII.  
 DR PRODOM; PD000131; Copper\_CUA; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 KW Copper; Oxidoreductase; Transmembrane.  
 SQ SEQUENCE 324 AA; 34739 MW; E16B6CC160899F72 CRC64;



DR PROSITE; PS00078; COX2; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 327 AA; 35006 MW; 3369DCBE31BA2E2 CRC64;

Query Match 79.3%; Score 23; DB 16; Length 327;  
 Best Local Similarity 37.5%; Pred. No. 1.8e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8

Db 126 HASAHVAH 133

## RESULT 9

Q8T468 PRELIMINARY; PRT; 341 AA.  
 AC Q8T468;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE AT13703P.  
 GN BCDNA:AT13703.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,  
 RA Miranda A., Murgall C.J., Nunco J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celnikier S.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY089327; AAL90065.1; -;  
 DR FlyBase; FBgn0063732; BCDNA:NT13703.  
 DR SQ SEQUENCE 341 AA; 34098 MW; EE6EBF27B2EC619A CRC64;

Query Match 79.3%; Score 23; DB 5; Length 341;  
 Best Local Similarity 37.5%; Pred. No. 1.9e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8

Db 287 HSHAAH 294

## RESULT 10

Q8MY62 PRELIMINARY; PRT; 341 AA.  
 AC Q8MY62;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Casein kinase I.  
 GN CS-CKI.  
 OS Ciona savignyi.  
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 OC Phlebobranchia; Cionidae; Ciona.  
 OC NCBI\_TaxID=51511;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Satou Y., Satoh N.;  
 RT "Ciona savignyi maternal gene.";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB078412; BAC05520.1; -;  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.

DR SMART; SM00220; S\_TKC; 1.  
 DR SMART; SM00219; TyrKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Kinase; Transferase.  
 SQ SEQUENCE 341 AA; 39427 MW; 9BA6B5348512C185 CRC64;

Query Match 79.3%; Score 23; DB 5; Length 341;  
 Best Local Similarity 37.5%; Pred. No. 1.9e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8

Db 321 HTQHSH 328

## RESULT 11

Q18888 PRELIMINARY; PRT; 342 AA.  
 ID Q18888  
 AC Q18888;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical 39.5 kDa protein.  
 GN C56C10.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Fulton L.;  
 RT "The sequence of C. elegans cosmid C56C10.";  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U29488; AAA68778.1; -;  
 DR WormPep; C56C10.10; CE02564.  
 DR InterPro; IPR001440; TPR.  
 DR Pfam; PF00515; TPR; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 342 AA; 39470 MW; 85EFE2A77F8D766B CRC64;

Query Match 79.3%; Score 23; DB 5; Length 342;  
 Best Local Similarity 37.5%; Pred. No. 1.9e+03;  
 Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8

Db 128 HSHAHTH 135

## RESULT 12

O54173 PRELIMINARY; PRT; 382 AA.  
 ID O54173  
 AC O54173;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein SCO5944.  
GN SCO5944 OR SC7H1.14.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3 (2) / M145;  
RX MEDLINE=21996410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
Harper D., Batenan A., Brown S., Chandra G., Chen C.W., Collins M.,  
Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
Hopwood D.A.;  
RA "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL939125; CAAL6201.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 382 AA; 41006 MW; 17C54D56069CE871 CRC64;  
  
Query Match 79.3%; Score 23; DB 16; Length 382;  
Best Local Similarity 37.5%; Pred. No. 2.1e+03;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 HXXXHXXH 8  
Db 370 HAAHAAH 377  
  
RESULT 13  
Q9KX27 PRELIMINARY; PRT; 402 AA.  
ID Q9KX27;  
AC Q9KX27; (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE CoxC protein.  
GN COXC.  
OS Oligotropha carboxydovorans (Pseudomonas carboxydovorans).  
OG Plasmid pHCG3.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Oligotropha.  
OX NCBI\_TaxID=40137;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OM5;  
RX Moersdorf G.;  
RA Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OM5;  
RX Schuebel U.;  
RA Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OM5;  
RX MEDLINE=95238294; PubMed=7721710;  
RA Schuebel U., Kraut M., Moersdorf G., Meyer O.;  
RT "Molecular Characterization of the Gene Cluster encoding the  
Molybdenum-Containing Carbon Monoxide Dehydrogenase of Oligotropha  
carboxydovorans.";  
RL J. Bacteriol. 177:2197-2197(1995).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OM5;  
RX Santiago B., Schuebel U., Egelseer C., Meyer O.;  
RT "Sequence analysis, characterization and CO-specific transcription of

RT the cox gene cluster on the megaplasmid pHCG3 of Oligotropha  
carboxydovorans.";  
RL Gene 236:1157-1247(1999).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OM5;  
RX Santiago B.;  
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RL EMBL; X82447; CAB76246.1; --  
DR InterPro; IPR005330; SPNT\_Repeat.  
DR Pfam; PF03707; MHT; 4.  
KW Plasmid.  
SQ SEQUENCE 402 AA; 42864 MW; 4C6108F085CA133D CRC64;  
  
Query Match 79.3%; Score 23; DB 2; Length 402;  
Best Local Similarity 37.5%; Pred. No. 2.1e+03;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 HXXXHXXH 8  
Db 127 HASAETH 134  
  
RESULT 14  
Q98UI5 PRELIMINARY; PRT; 412 AA.  
ID Q98UI5;  
AC Q98UI5; (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Thyroid transcription factor-1.  
GN LJTTF-1.  
OS Lampetra japonica (Japanese lamprey) (Eutrophenus japonicus).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;  
OC Petromyzontiformes; Petromyzontidae; Lethenteron.  
OX NCBI\_TaxID=94989;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ventral forebrain;  
RA Ogasawara M., Shigetani Y., Suzuki S., Kuratani S., Satoh N.;  
RT "Expression of Thyroid Transcription Factor-1 (TTF-1) Gene in the  
Ventral Forebrain and Endostyle of the Agnathan Vertebrate, Lampetra  
japonica.";  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
DR EMBL; AB052339; BAB32434.1; --  
DR HSSP; P23441; 1FTT.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00024; HOMEBOX.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW DNA-binding; Homeobox; Nuclear protein.  
SQ SEQUENCE 412 AA; 43509 MW; EC844185CD89D5EB CRC64;  
  
Query Match 79.3%; Score 23; DB 13; Length 412;  
Best Local Similarity 37.5%; Pred. No. 2.2e+03;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 HXXXHXXH 8  
Db 114 HHAHAAH 121  
  
RESULT 15  
Q947K4 PRELIMINARY; PRT; 466 AA.  
ID Q947K4;  
AC Q947K4;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE Thiohydroximate S-glucosyltransferase.  
OS Brassica napus (Rape).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Brassica.  
CX NCBI\_TaxID=3708;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Marilia E.-F.A., MacPherson J.M., Tsang E.W.T., Van Audenhove K.,  
RA Keller W.A., Grootwassink J.W.D.;  
RT "Molecular cloning of a Brassica napus thiohydroximate S-  
RT glucosyltransferase gene and its expression in Escherichia coli.";  
RL Physiol. Plantarum 0:0-0(2001).  
DR EMBL; AF304430; AAL09350.1; -;  
DR InterPro; IPR002213; UDP-glucosyl-trans.  
DR Pfam; PF00201; UDPGT; 1.  
DR PROSITE; PS00375; UDPGT; 1.  
KW Transferase.  
SQ SEQUENCE 466 AA; 50826 MW; D5991B82129C2C1C CRC64;

Query Match 79.3%; Score 23; DB 10; Length 466;  
Best Local Similarity 37.5%; Pred.No. 2.4e+03;  
Matches 3; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 HXXXHXXH 8  
Db 188 HSSSHAEH 195

Search completed: November 12, 2003, 09:36:48  
Job time : 30 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:31:02 ; Search time 35.5 Seconds  
(without alignments)  
35.769 Million cell updates/sec

Title: US-10-064-903-2

Perfect score: 46

Sequence: 1 HDLHLVLH 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 19Jun03.\*

```
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	91.3	11	24	AAE29884
2	42	91.3	439	23	ABG69073
3	42	91.3	441	23	ABG69068
4	42	91.3	441	23	ABG69076
5	42	91.3	441	23	ABB86654
6	42	91.3	444	23	ABG69086
7	42	91.3	548	19	AAW56014
8	42	91.3	848	23	ABG69087
9	42	91.3	852	23	ABG69077
10	42	91.3	858	19	AAW56018
11	42	91.3	1169	19	AAW56017
12	42	91.3	1291	19	AAW68392
13	42	91.3	1315	22	AAW61169
14	38	82.6	436	23	ABG69072
15	38	82.6	443	23	ABG69084
16	38	82.6	858	23	ABG69085
17	38	82.6	907	20	AAW34888
18	36	78.3	104	23	ABP34326
19	36	78.3	422	23	ABG69071
20	36	78.3	427	23	ABG69082
21	36	78.3	489	23	ABB55296
22	36	78.3	773	23	ABB91782
23	36	78.3	804	23	ABG69083
24	35	76.1	60	15	AAW50716
25	35	76.1	60	17	AAW02908
26	35	76.1	155	23	ABU51691
27	35	76.1	240	23	ABU51965
28	35	76.1	251	22	ABB10271
29	35	76.1	251	22	AAU22991
30	35	76.1	251	23	ABP66858
31	35	76.1	269	14	AAW44805
32	35	76.1	269	24	ABU03642
33	35	76.1	315	18	AAW20813
34	35	76.1	737	24	ABU00661
35	34	73.9	9	22	AAW98985
36	34	73.9	118	22	AAW63297
37	34	73.9	119	23	ABP34926
38	34	73.9	150	22	AAW63290
39	34	73.9	182	22	AAU37971
40	34	73.9	198	23	ABW47377
41	34	73.9	198	24	ABU02143
42	34	73.9	226	19	AAW85978
43	34	73.9	452	22	AAU38112
44	34	73.9	456	22	AAU34657
45	34	73.9	498	22	ABG25738

#### ALIGNMENTS

RESULT 1  
AAE29884  
ID AAE29884 standard; peptide; 11 AA.  
XX  
AC AAE29884;  
XX  
DT 24-FEB-2003 (first entry)  
XX  
DE Clostridium tetani TeTx L chain fragment.  
XX  
KW Gonadotropin releasing hormone analogue; neurotoxin; prostate cancer;  
KW endocrine disorder; gonadotropin related illness; endometrial cancer;  
KW pancreatic cancer; breast cancer; endometriosis; precocious puberty;  
KW GnRH-A; therapy; protease; L chain; tetani toxin; TeTx.  
XX  
OS Clostridium tetani.  
XX  
PN WO200274327-A2.  
XX  
PD 26-SEP-2002.  
XX  
PF 11-MAR-2002; 2002WO-US07379.  
XX  
PR 15-MAR-2001; 2001US-0810601.  
XX  
PA (ALLR ) ALLERGAN SALES INC.  
XX  
PI Donovan S;  
XX  
DR WPI; 2003-018772/01.  
XX  
PT New agent comprising a light chain and a (modified) heavy chain of a

Recombinant botuli  
Recombinant botuli  
Clostridium botuli  
Clostridium tetani  
Botulinum neurotox  
Botulinum neurotox  
Botulinum neurotox  
Amino acid sequenc  
Human ORF3299 prot  
Botulinum neurotox  
Botulinum neurotox  
Lactococcus lactis  
Herbicidally activ  
Botulinum neurotox  
G-protein coupled  
G-protein coupled  
Helicobacter pylori  
Helicobacter pylori  
Human cDNA 58Q ID  
Novel human enzyme  
Human polypeptide  
Human cyclin D2 ps  
Human expressed pr  
S. pneumoniae type  
H. pylori secreted  
Vaccine related MH  
Human breast cancer  
Human ORF3899 prot  
Human breast cancer  
Streptococcus pneu  
Listeria monocytog  
S. pneumoniae type  
S. pneumoniae type  
S. pneumoniae dari  
Salmonella typhi c  
E. coli cellular p  
Novel human diagno



CC to control autonomic nerve function or tiptoe-walking due to stiff  
 CC muscles common in children with cerebral palsy. The sequences are also  
 CC useful for screening for botulinum neurotoxin inhibitors. This sequence  
 CC represents a botulinum neurotoxin light chain serotype A protein.

SQ Sequence 441 AA;  
 Query Match 91.3%; Score 42; DB 23; Length 441;  
 Best Local Similarity 87.5%; Pred. No. 13;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8  
 :|||||  
 Db 230 HELIHVLH 237

RESULT 4  
 ABG69076  
 ID ABG69076 standard; Protein; 441 AA.  
 XX AC ABG69076;

DT 07-OCT-2002 (first entry)

DE Botulinum neurotoxin light chain polypeptide #10.

XX Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;  
 KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;  
 KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;  
 KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;  
 KW lower motor neuron hyperactivity; autonomic nerve function; muscular;  
 KW immunostimulant; antibacterial.

XX Clostridium botulinum.

OS WO200236758-A2.

PN 10-MAY-2002.

XX 06-NOV-2001; 2001WO-US47230.

XX 06-NOV-2000; 2000US-246774P.

PR 20-JUL-2001; 2001US-0910186.

PR 09-AUG-2001; 2001US-311966P.

XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

PA Smith LA, Jensen M;

PI WPI; 2002-575192/61.

DR N-PSDB; ABK98546.

XX Novel nucleic acid molecule encoding botulinum neurotoxin light chain  
 PT serotype A, useful for producing the neurotoxin for vaccination against  
 PT botulism, comprises sequence expressible in host other than Clostridium  
 PT -  
 XX Claim 33; Page 135-136; 166pp; English.

XX The invention relates to a nucleic acid molecule encoding a botulinum  
 CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence  
 CC that is expressible in a host organism other than Clostridium, or has a  
 CC total A+T content that is less than about 70%. The BoNT LC protein is  
 CC useful in vaccination against botulism, for eliciting protective immunity  
 CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,  
 CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental  
 CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,  
 CC conditions characterised by hyperactivity of the lower motor neuron, and  
 CC to control autonomic nerve function or tiptoe-walking due to stiff  
 CC muscles common in children with cerebral palsy. The sequences are also  
 CC useful for screening for botulinum neurotoxin inhibitors. This sequence  
 CC represents a botulinum neurotoxin light chain serotype A protein.

XX

SQ Sequence 441 AA;

Query Match 91.3%; Score 42; DB 23; Length 441;  
 Best Local Similarity 87.5%; Pred. No. 13;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8  
 :|||||  
 Db 230 HELIHVLH 237

RESULT 5  
 ABB80654  
 ID ABB80654 standard; peptide; 441 AA.

XX AC ABB80654;

XX 15-JUL-2002 (first entry)

DE Botulinum toxin type B Danish I light chain.

XX Neurotoxin; biological persistence; dysphonia; strabismus; muscle spasm;  
 KW dystonia; pain; blepharospasm; hemifacial spasm; excessive salivation;  
 KW eyelid disorder; cerebral palsy; focal spasticity; spasmodic colitis;  
 KW neurogenic bladder; anismus; limb spasticity; tic; tremor; bruxism;  
 KW anal fissure; achalasia; dysphagia; lacrimation; hyperhidrosis; headache;  
 KW excessive gastrointestinal secretion; botulinum toxin type B; Danish I;  
 KW light chain.

XX Clostridium botulinum.

OS WO200208268-A2.

PN 31-JAN-2002.

XX 20-JUL-2001; 2001WO-US23122.

XX 21-JUL-2000; 2000US-0620840.

XX (ALLR ) ALLERGAN SALES INC.

PI Steward LE, Fernandez-salas E, Herrington TW, Aoki KR;

DR WPI; 2002-241566/29.

XX Novel modified neurotoxin comprising structural modification which  
 PT alters the biological persistence and/or biological activity of a  
 PT neurotoxin, useful for treating neuromuscular or autonomic disorder, or  
 PT pain -  
 XX Disclosure; Fig 8; 102pp; English.

XX The sequence represents the botulinum toxin type B Danish I light chain.  
 CC The invention relates to a novel modified neurotoxin including a  
 CC structural modification, where the structural modification is effective  
 CC to alter the biological persistence, or biological activity. The modified  
 CC neurotoxin is useful for treating spasmodic dysphonia, laryngeal  
 CC dystonia, oromandibular dysphonia, lingual dystonia, cervical dystonia,  
 CC focal hand dystonia, blepharospasm, strabismus, hemifacial spasm, eyelid  
 CC disorder, cerebral palsy, focal spasticity, spasmodic colitis, neurogenic  
 CC bladder, anismus, limb spasticity, tics, tremors, bruxism, anal fissure,  
 CC achalasia, dysphagia, lacrimation, hyperhidrosis, excessive salivation,  
 CC excessive gastrointestinal secretions, pain from muscle spasms, headache  
 CC pain, brow furrows or skin wrinkles.

XX SQ Sequence 441 AA;

Query Match 91.3%; Score 42; DB 23; Length 441;  
 Best Local Similarity 87.5%; Pred. No. 13;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8  
 :|||||

Db 230 HELIHLVH 237

RESULT 6

ABG69086

ID ABG69086 standard; Protein; 444 AA.

XX AC ABG69086;

XX DT 07-OCT-2002 (first entry)

XX DE Botulinum neurotoxin light chain polypeptide #20.

XX KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;

XX KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;

XX KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;

XX KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;

XX KW lower motor neuron hyperactivity; autonomic nerve function; muscular;

XX KW immunostimulant; antibacterial.

XX OS Clostridium botulinum.

XX PN WO200236758-A2.

XX PN 10-MAY-2002.

XX PF 06-NOV-2001; 2001WO-US47230.

XX PR 06-NOV-2000; 2000US-246774P.

XX PR 20-JUL-2001; 2001US-0910186.

XX PR 09-AUG-2001; 2001US-311966P.

XX PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

XX PI Smith LA, Jensen M;

XX WPI; 2002-575192/61.

XX N-PSDB; ABK98556.

XX PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain

PT serotype A, useful for producing the neurotoxin for vaccination against

PT botulism, comprises sequence expressible in host other than Clostridium

PT -

XX Claim 33; Page 160-161; 166pp; English.

XX The invention relates to a nucleic acid molecule encoding a botulinum

CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence

CC that is expressible in a host organism other than Clostridium, or has a

CC total A+T content that is less than about 70% The BoNT LC protein is

CC useful in vaccination against botulism, for eliciting protective immunity

CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,

CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental

CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,

CC conditions characterised by hyperactivity of the lower motor neuron, and

CC to control autonomic nerve function or tiptoe-walking due to stiff

CC muscles common in children with cerebral palsy. The sequences are also

CC useful for screening for botulinum neurotoxin inhibitors. This sequence

CC represents a botulinum neurotoxin light chain serotype A protein.

XX Sequence 444 AA;

Query Match 91.3%; Score 42; DB 23; Length 444;

Best Local Similarity 87.5%; Pred. No. 13;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHLVH 8

Db 230 HELIHLVH 237

RESULT 7

ABG69014

ID ABG69014 standard; Protein; 548 AA.

XX AC AAW56014;

XX DT 27-JUL-1998 (first entry)

XX DE Recombinant botulinum neurotoxin type B LH107/B.

XX KW Botulinum; recombinant; Clostridium botulinum; neurotoxin;

XX KW immunogen; detection; tetanus; non-toxic; toxin.

XX OS Synthetic.

XX OS Clostridium botulinum.

XX PN WO9807864-A1.

XX PD 26-FEB-1998.

XX PF 22-AUG-1997; 97WO-GB02273.

XX PR 13-DEC-1996; 96GB-0025996.

XX PR 23-AUG-1996; 96GB-0017671.

XX PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.

PA (SPEY-) SPEYWOOD LAB LTD.

XX PI Foster KA, Quinn CP, Shone CC;

XX WPI; 1998-169168/15.

XX N-PSDB; AAV26286.

XX PT Recombinant neurotoxin polypeptides - used to develop therapeutic

PT agents, immunogens or as non-toxic standards for the detection of

PT neurotoxins

XX Example 2; Page 103-104; 137pp; English.

XX The present sequence represents a recombinant neurotoxin protein from

CC the present invention. The present invention describes recombinant

CC neurotoxin proteins which comprise a first and second domain, where

CC the first domain is adapted to cleave one or more vesicle or

CC plasma-membrane associated proteins essential to exocytosis, and where

CC the second domain is adapted: (a) to translocate the protein into a

CC cell; (b) to increase the solubility of the protein compared to the

CC solubility of the first domain on its own, or (c) both to translocate

CC the protein into a cell and to increase the solubility of the protein

CC compared to the solubility of the first domain on its own, the protein

CC being free of clostridial neurotoxin (CN) and free of CN precursor that

CC proteins can be used as therapeutic agents for targeting cells

CC expressing a relevant substrate. The products can also be used as

CC immunogens and as non-toxic standards for the assessment and development

CC of in vitro assays for the detection of functional botulinum or tetanus

CC neurotoxins either in foodstuffs or in environmental samples.

XX Sequence 548 AA;

Query Match 91.3%; Score 42; DB 19; Length 548;

Best Local Similarity 87.5%; Pred. No. 16;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHLVH 8

Db 230 HELIHLVH 237

RESULT 8

ABG69087

ID ABG69087 standard; Protein; 848 AA.

XX AC ABG69087;

XX DT 07-OCT-2002 (first entry)

XX Botulinum neurotoxin light chain polypeptide #21.  
 XX Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;  
 KW spasticity; ocular motility; facial dyskinesias; stiff-person syndrome;  
 KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;  
 KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;  
 KW lower motor neuron hyperactivity; autonomic nerve function; muscular;  
 KW immunostimulant; antibacterial.  
 XX Clostridium botulinum.  
 OS Clostridium botulinum.  
 PN WO200236758-A2.  
 XX 10-MAY-2002.  
 XX 06-NOV-2001; 2001WO-US47230.  
 XX 06-NOV-2000; 2000US-246774P.  
 PR 20-JUL-2001; 2001US-0910186.  
 PR 09-AUG-2001; 2001US-311966P.  
 XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
 PA Smith LA, Jensen M;  
 PI WPI; 2002-575192/61.  
 DR N-PSDB; ABK98557.  
 XX Novel nucleic acid molecule encoding botulinum neurotoxin light chain  
 PT serotype A, useful for producing the neurotoxin for vaccination against  
 PT botulism, comprises sequence expressible in host other than Clostridium  
 PT -  
 PS Claim 52; Page 162-164; 166pp; English.  
 CC The invention relates to a nucleic acid molecule encoding a botulinum  
 CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence  
 CC that is expressible in a host organism other than Clostridium, or has a  
 CC total A+T content that is less than about 70% The BoNT LC protein is  
 CC useful in vaccination against botulism, for eliciting protective immunity  
 CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,  
 CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental  
 CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,  
 CC conditions characterised by hyperactivity of the lower motor neuron, and  
 CC to control autonomic nerve function or tiptoe-walking due to stiff  
 CC muscles common in children with cerebral palsy. The sequences are also  
 CC useful for screening for botulinum neurotoxin inhibitors. This sequence  
 CC represents a botulinum neurotoxin light chain serotype A protein.  
 XX Sequence 848 AA;  
 SQ Query Match 91.3%; Score 42; DB 23; Length 848;  
 Best Local Similarity 87.5%; Pred. No. 25;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HDLIHVLH 8  
 Db :|||||  
 230 HELIHVLH 237  
 RESULT 9  
 ABG69077  
 ID ABG69077 standard; Protein; 852 AA.  
 XX AC  
 XX ABG69077;  
 DT 07-OCT-2002 (first entry)  
 XX DE Botulinum neurotoxin light chain polypeptide #11.  
 XX Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;  
 KW spasticity; ocular motility; facial dyskinesias; stiff-person syndrome;  
 KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;  
 KW spasticity; ocular motility; facial dyskinesias; stiff-person syndrome;

KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;  
 KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;  
 KW lower motor neuron hyperactivity; autonomic nerve function; muscular;  
 KW immunostimulant; antibacterial.  
 XX Clostridium botulinum.  
 OS Clostridium botulinum.  
 PN WO200236758-A2.  
 XX 10-MAY-2002.  
 XX 06-NOV-2001; 2001WO-US47230.  
 XX 06-NOV-2000; 2000US-246774P.  
 PR 20-JUL-2001; 2001US-0910186.  
 PR 09-AUG-2001; 2001US-311966P.  
 XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
 PA Smith LA, Jensen M;  
 PI WPI; 2002-575192/61.  
 DR N-PSDB; ABK98547.  
 XX Novel nucleic acid molecule encoding botulinum neurotoxin light chain  
 PT serotype A, useful for producing the neurotoxin for vaccination against  
 PT botulism, comprises sequence expressible in host other than Clostridium  
 PT -  
 PS Claim 52; Page 138-139; 166pp; English.  
 CC The invention relates to a nucleic acid molecule encoding a botulinum  
 CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence  
 CC that is expressible in a host organism other than Clostridium, or has a  
 CC total A+T content that is less than about 70% The BoNT LC protein is  
 CC useful in vaccination against botulism, for eliciting protective immunity  
 CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,  
 CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental  
 CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,  
 CC conditions characterised by hyperactivity of the lower motor neuron, and  
 CC to control autonomic nerve function or tiptoe-walking due to stiff  
 CC muscles common in children with cerebral palsy. The sequences are also  
 CC useful for screening for botulinum neurotoxin inhibitors. This sequence  
 CC represents a botulinum neurotoxin light chain serotype A protein.  
 XX Sequence 852 AA;  
 SQ Query Match 91.3%; Score 42; DB 23; Length 852;  
 Best Local Similarity 87.5%; Pred. No. 26;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HDLIHVLH 8  
 Db :|||||  
 229 HELIHVLH 236  
 RESULT 10  
 AAW56018  
 ID AAW56018 standard; Protein; 858 AA.  
 XX AC  
 XX AAW56018;  
 DT 27-JUL-1998 (first entry)  
 XX DE Recombinant botulinum neurotoxin type B LH417/B.  
 XX Botulinum; recombinant; Clostridium botulinum; neurotoxin;  
 KW immunogen; detection; tetanus; non-toxic; toxin.  
 XX Synthetic.  
 OS Clostridium botulinum.  
 XX WO9807864-A1.



```

XX PD 26-FEB-1998.
XX PF 22-AUG-1997; 97WO-GB02273.
XX PR 13-DEC-1996; 96GB-0025996.
XX PR 23-AUG-1996; 96GB-0017671.
XX FA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX PA (SPEY-) SPEYWOOD LAB LTD.
XX PI Foster KA, Quinn CP, Shone CC;
XX DR WPI; 1998-169168/15.
XX DR N-PSDB; AAV26290.
XX PT Recombinant neurotoxin polypeptides - used to develop therapeutic
XX PT agents, immunogens or as non-toxic standards for the detection of
XX PT neurotoxins
XX PS Example 2; Page 98-100; 137pp; English.
XX CC The present sequence represents a recombinant neurotoxin protein from
XX CC the present invention. The present invention describes recombinant
XX CC neurotoxin proteins which comprise a first and second domain, where
XX CC the first domain is adapted to cleave one or more vesicle or
XX CC plasma-membrane associated proteins essential to exocytosis, and where
XX CC the second domain is adapted: (a) to translocate the protein into a
XX CC cell; (b) to increase the solubility of the protein compared to the
XX CC solubility of the first domain on its own, or (c) both to translocate
XX CC the protein into a cell and to increase the solubility of the protein
XX CC compared to the solubility of the first domain on its own, the protein
XX CC being free of clostridial neurotoxin (CN) and free of CN precursor that
XX CC can be converted into toxin by proteolytic action. The recombinant
XX CC proteins can be used as therapeutic agents for targeting cells
XX CC expressing a relevant substrate. The products can also be used as
XX CC immunogens and as non-toxic standards for the assessment and development
XX CC of in vitro assays for the detection of functional botulinum or tetanus
XX CC neurotoxins either in foodstuffs or in environmental samples.
XX SQ Sequence 858 AA;
Query Match 91.3%; Score 42; DB 19; Length 858;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
Db 230 HELIHLVLH 237
RESULT 11
AAW56017
ID AAW56017 standard; Protein; 1169 AA.
XX AC AAW56017;
XX DT 27-JUL-1998 (first entry)
XX DE Recombinant botulinum neurotoxin type B LH728/B.
XX KW Botulinum; recombinant; Clostridium botulinum; neurotoxin;
XX KW immunogen; detection; tetanus; non-toxic; toxin.
XX OS Synthetic.
XX OS Clostridium botulinum.
XX PN WO9807864-A1.
XX PR 26-FEB-1998.
XX PD 22-AUG-1997; 97WO-GB02273.
XX PF Thalley BS, Williams JA;

```

```

PR 13-DEC-1996; 96GB-0025996.
PR 23-AUG-1996; 96GB-0017671.
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX PA (SPEY-) SPEYWOOD LAB LTD.
XX PI Foster KA, Quinn CP, Shone CC;
XX DR WPI; 1998-169168/15.
XX DR N-PSDB; AAV26289.
XX PT Recombinant neurotoxin polypeptides - used to develop therapeutic
XX PT agents, immunogens or as non-toxic standards for the detection of
XX PT neurotoxins
XX PS Example 2; Page 91-94; 137pp; English.
XX CC The present sequence represents a recombinant neurotoxin protein from
XX CC the present invention. The present invention describes recombinant
XX CC neurotoxin proteins which comprise a first and second domain, where
XX CC the first domain is adapted to cleave one or more vesicle or
XX CC plasma-membrane associated proteins essential to exocytosis, and where
XX CC the second domain is adapted: (a) to translocate the protein into a
XX CC cell; (b) to increase the solubility of the protein compared to the
XX CC solubility of the first domain on its own, or (c) both to translocate
XX CC the protein into a cell and to increase the solubility of the protein
XX CC compared to the solubility of the first domain on its own, the protein
XX CC being free of clostridial neurotoxin (CN) and free of CN precursor that
XX CC can be converted into toxin by proteolytic action. The recombinant
XX CC proteins can be used as therapeutic agents for targeting cells
XX CC expressing a relevant substrate. The products can also be used as
XX CC immunogens and as non-toxic standards for the assessment and development
XX CC of in vitro assays for the detection of functional botulinum or tetanus
XX CC neurotoxins either in foodstuffs or in environmental samples.
XX SQ Sequence 1169 AA;
Query Match 91.3%; Score 42; DB 19; Length 1169;
Best Local Similarity 87.5%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HDLIHVLH 8
Db 230 HELIHLVLH 237
RESULT 12
AAW68392
ID AAW68392 standard; Protein; 1291 AA.
XX AC AAW68392;
XX DT 07-DEC-1998 (first entry)
XX DE Clostridium botulinum type B toxin.
XX KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
XX KW botulism; BotB.
XX OS Clostridium botulinum serotype B Danish strain.
XX PN WO9808540-A1.
XX PR 05-MAR-1998.
XX PD 28-AUG-1997; 97WO-US15394.
XX PF 28-AUG-1996; 96US-0704159.
XX PR 28-AUG-1996; 96US-0704159.
XX FA (OPHI-) OPHIDIAN PHARM INC.
XX PI Thalley BS, Williams JA;

```

DR WPI: 1998-230234/20.  
 DR N-PSDB; AAV30579.  
 XX Host cell containing recombinant expression vector encoding  
 PT Clostridium botulinum type B or E toxin - useful to treat humans  
 PT and other animals at risk of intoxication with clostridial toxin  
 XX  
 PS Example 35; Page 291-296; 428pp; English.  
 XX  
 CC This is the amino acid sequence of the type B toxin of Clostridium  
 CC botulinum Danish strain. The C fragment (see AA68393-94) of the  
 CC serotype B toxin has been expressed as a histidine-tagged protein  
 CC in Escherichia coli. The invention relates to C. botulinum  
 CC recombinant toxins. Methods are provided which allow for the  
 CC isolation of soluble recombinant proteins free of significant  
 CC endotoxin contamination. Preferred hosts for production of the  
 CC recombinant proteins are E. coli, insect cells and yeast cells.  
 CC The recombinant proteins are used as immunogens for the production  
 CC of vaccines and antitoxins that are useful in the treatment of  
 CC humans and animals at risk of intoxication with clostridial toxin.  
 XX  
 SQ Sequence 1291 AA;  
 Query Match 91.3%; Score 42; DB 19; Length 1291;  
 Best Local Similarity 87.5%; Pred. No. 39;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HDLIHVLH 8  
 DB 230 HELIHVLH 237  
 [|||||]  
 RESULT 13  
 RAB61169  
 ID AAB61169 standard; Protein; 1315 AA.  
 AC AAB61169;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Clostridium tetani TeNT.  
 XX  
 KW Clostridium tetani; TeNT; tetanus toxin; antibacterial; vaccine;  
 KW TeNT fragment C; infection.  
 KW  
 XX Clostridium tetani.  
 OS  
 XX WO200100839-A1.  
 XX  
 XX 04-JAN-2001.  
 XX  
 XX 23-JUN-2000; 2000WO-GB02428.  
 XX  
 XX 25-JUN-1999; 99GB-0014861.  
 XX  
 XX (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.  
 XX  
 PI Fairweather NP, Sinha K;  
 XX  
 XX WPI: 2001-123014/13.  
 XX  
 PT New polypeptide, useful for treating infections of Clostridium tetani,  
 PT comprises tetanus toxin fragment with a mutation in a loop region, -  
 PT  
 PS Disclosure; Page 39; 43pp; English.  
 XX  
 CC The present sequence is given in a specification relating to a novel  
 CC polypeptide comprising tetanus toxin (TeNT) fragment C or its immunogenic  
 CC fragment, containing a mutation in a loop region. The mutation results in  
 CC a reduction in the binding of TeNT fragment C or its immunogenic fragment  
 CC to gangliosides and primary motoneurons, and/or a reduction in the  
 CC ability of TeNT fragment C or its immunogenic fragment to undergo  
 CC retrograde transport. The polypeptide is useful for treating, preventing

CC and reducing the susceptibility to Clostridium tetani infection in a  
 CC human or animal, and also for producing antibodies which recognise groups  
 CC within TeNT polypeptides. Antibody produced against the polypeptide is  
 CC also useful for treating Clostridium tetani infection.  
 XX  
 SQ Sequence 1315 AA;  
 Query Match 91.3%; Score 42; DB 22; Length 1315;  
 Best Local Similarity 87.5%; Pred. No. 40;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HDLIHVLH 8  
 DB 233 HELIHVLH 240  
 [|||||]  
 RESULT 14  
 ABG69072  
 ID ABG69072 standard; Protein; 436 AA.  
 XX  
 AC ABG69072;  
 XX  
 DT 07-OCT-2002 (first entry)  
 XX  
 DE Botulinum neurotoxin light chain polypeptide #6.  
 XX  
 KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;  
 KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;  
 KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;  
 KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;  
 KW lower motor neuron hyperactivity; autonomic nerve function; muscular;  
 KW immunostimulant; antibacterial.  
 XX  
 XX Clostridium botulinum.  
 OS  
 XX WO200236758-A2.  
 XX  
 XX 10-MAY-2002.  
 XX  
 XX 06-NOV-2001; 2001WO-US47230.  
 XX  
 XX 06-NOV-2000; 2000US-246774P.  
 PR 20-JUL-2001; 2001US-0910186.  
 PR 09-AUG-2001; 2001US-311966P.  
 XX  
 XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
 XX  
 XX Smith LA, Jensen M;  
 PI  
 XX WPI: 2002-575192/61.  
 DR N-PSDB; ABK98542.  
 XX  
 PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain  
 PT serotype A, useful for producing the neurotoxin for vaccination against  
 PT botulism, comprises sequence expressible in host other than Clostridium  
 PT  
 PS Claim 33; Page 127-128; 166pp; English.  
 XX  
 CC The invention relates to a nucleic acid molecule encoding a botulinum  
 CC neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence  
 CC that is expressible in a host organism other than Clostridium, or has a  
 CC total A+T content that is less than about 70% The BoNT LC protein is  
 CC useful in vaccination against botulism, for eliciting protective immunity  
 CC in a mammal, for treating dystonias, spasticity, pain, ocular motility,  
 CC facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental  
 CC myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,  
 CC conditions characterised by hyperactivity of the lower motor neuron, and  
 CC to control autonomic nerve function or tiptoe-walking due to stiff  
 CC muscles common in children with cerebral palsy. The sequences are also  
 CC useful for screening for botulinum neurotoxin inhibitors. This sequence  
 CC represents a botulinum neurotoxin light chain serotype A protein.  
 XX

SQ Sequence 436 AA;

Query Match 82.6%; Score 38; DB 23; Length 436;  
Best Local Similarity 75.0%; Pred. No. 67;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
: : : : :  
Db 225 HELIHALH 232

# RESULT 15

ABG69084  
ID ABG69084 standard; Protein; 443 AA.

XX AC ABG69084;

DT 07-OCT-2002 (first entry)

DE Botulinum neurotoxin light chain polypeptide #18.

KW Botulinum neurotoxin light chain; BoNT LC; botulism; dystonia; pain;  
KW spasticity; ocular motility; facial dyskinesia; stiff-person syndrome;  
KW bladder dysfunction; segmental myoclonus; hyperkinetic disorder;  
KW cosmetic treatment; facial wrinkle; cerebral palsy; analgesic; relaxant;  
KW lower motor neuron hyperactivity; autonomic nerve function; muscular;  
KW immunostimulant; antibacterial.

XX OS Clostridium botulinum.

XX FN WO200236758-A2.

XX PD 10-MAY-2002.

XX PF 06-NOV-2001; 2001WO-US47230.

XX PR 06-NOV-2000; 2000US-246774P.

XX PR 20-JUL-2001; 2001US-0910186.

XX PR 09-AUG-2001; 2001US-311966P.

XX PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

XX FI Smith LA, Jensen M;

XX DR WPI; 2002-575192/61.

XX DR N-PSDB; ABK98554.

XX PT Novel nucleic acid molecule encoding botulinum neurotoxin light chain  
serotype A, useful for producing the neurotoxin for vaccination against  
botulism, comprises sequence expressible in host other than Clostridium

PS Claim 33; Page 155-156; 166pp; English.

XX The invention relates to a nucleic acid molecule encoding a botulinum  
neurotoxin light chain (BoNT LC) serotype A, where the DNA has a sequence  
that is expressible in a host organism other than Clostridium, or has a  
total A+T content that is less than about 70% The BoNT LC protein is  
useful in vaccination against botulism, for eliciting protective immunity  
in a mammal, for treating dystonias, spasticity, pain, ocular motility,  
facial dyskinesias, stiff-person syndrome, bladder dysfunction, segmental  
myoclonus, hyperkinetic disorders, cosmetic treatment of facial wrinkles,  
conditions characterised by hyperactivity of the lower motor neuron, and  
to control autonomic nerve function or tiptoe-walking due to stiff  
muscles common in children with cerebral palsy. The sequences are also  
useful for screening for botulinum neurotoxin inhibitors. This sequence  
represents a botulinum neurotoxin light chain serotype A protein.

XX SQ Sequence 443 AA;

Query Match 82.6%; Score 38; DB 23; Length 443;  
Best Local Similarity 75.0%; Pred. No. 67;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
: : : : :  
Db 227 HELIHALH 234

Search completed: November 12, 2003, 09:35:07  
Job time : 36.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:33:48 ; Search time 14 Seconds  
(without alignments)  
24.178 Million cell updates/sec

Title: US-10-064-903-2

Perfect score: 46

Sequence: 1 HDLHVH 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*
- 1: /cgn2\_6/ptodata/1/iaa/5A\_COVB.pep:\*
  - 2: /cgn2\_6/ptodata/1/iaa/5B\_COVB.pep:\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A\_COVB.pep:\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B\_COVB.pep:\*
  - 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COVB.pep:\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	91.3	548	4	US-09-255-829-24
2	42	91.3	858	4	US-09-255-829-22
3	42	91.3	858	4	US-09-255-829-29
4	42	91.3	1169	4	US-09-255-829-20
5	42	91.3	1315	4	US-08-913-880C-1
6	38	82.6	907	4	US-09-198-452A-306
7	35	76.1	60	1	US-08-118-270-257
8	35	76.1	60	5	PCT-US93-08528-257
9	35	76.1	269	9	PCT-US93-05000-31
10	34	73.9	204	4	US-09-107-532A-7103
11	33	71.7	60	1	US-08-117-083-20
12	33	71.7	65	1	US-07-879-685B-1
13	33	71.7	162	1	US-07-879-685B-4
14	33	71.7	416	1	US-08-117-083-62
15	33	71.7	431	1	US-08-311-023-2
16	33	71.7	541	2	US-08-540-804-16
17	33	71.7	541	2	US-08-218-265-16
18	33	71.7	541	3	US-08-521-872-16
19	33	71.7	541	3	US-08-590-399-16
20	33	71.7	942	4	US-09-657-931A-9
21	32	69.6	73	4	US-09-328-352-5131
22	32	69.6	285	4	US-09-328-352-7570
23	32	69.6	289	3	US-09-105-697-2
24	32	69.6	300	4	US-09-585-858-42
25	32	69.6	310	2	US-08-484-956-88
26	32	69.6	310	2	US-08-757-653-88
27	32	69.6	310	4	US-08-520-946-88

28	32	69.6	315	2	US-08-484-956-91	Sequence 91, Appl
29	32	69.6	315	2	US-08-757-653-91	Sequence 91, Appl
30	32	69.6	315	4	US-08-520-946-91	Sequence 91, Appl
31	32	69.6	320	2	US-08-757-653-163	Sequence 163, App
32	32	69.6	320	2	US-08-823-516-61	Sequence 61, Appl
33	32	69.6	320	3	US-08-759-038-102	Sequence 102, App
34	32	69.6	320	3	US-08-758-314-102	Sequence 102, App
35	32	69.6	320	4	US-09-684-938-102	Sequence 102, App
36	32	69.6	320	4	US-09-308-825A-102	Sequence 102, App
37	32	69.6	322	2	US-08-484-956-89	Sequence 89, Appl
38	32	69.6	322	2	US-08-757-653-89	Sequence 89, Appl
39	32	69.6	322	4	US-08-520-946-89	Sequence 89, Appl
40	32	69.6	359	4	US-09-252-991A-18134	Sequence 18134, A
41	32	69.6	528	2	US-08-484-956-90	Sequence 90, Appl
42	32	69.6	528	2	US-08-757-653-90	Sequence 90, Appl
43	32	69.6	528	4	US-08-520-946-90	Sequence 90, Appl
44	32	69.6	548	2	US-08-484-956-86	Sequence 86, Appl
45	32	69.6	548	2	US-08-757-653-86	Sequence 86, Appl

ALIGNMENTS

RESULT 1  
US-09-255-829-24  
; Sequence 24, Application US/09255829  
; Patent No. 6461617  
; GENERAL INFORMATION:  
; APPLICANT: Shone, Clifford Charles  
; APPLICANT: Quinn, Conrad Padraig  
; APPLICANT: Foster, Keith Alan  
; TITLE OF INVENTION: Recombinant Toxin Fragments  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.  
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/255,829  
; FILING DATE: 23-FEB-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB97/02273  
; FILING DATE: 22-AUG-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/782,893  
; FILING DATE: 27-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ESMOND, ROBERT W.  
; REGISTRATION NUMBER: 32,893  
; REFERENCE/DOCKET NUMBER: 1581.0130002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 548 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-255-829-24

Query Match 91.3%; Score 42; DB 4; Length 548;  
Best Local Similarity 87.5%; Pred No. 8.8;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 HDLHVH 8
Db      230 HELHVH 237

RESULT 2
US-09-255-829-22
; Sequence 22, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-22

Query Match      91.3%; Score 42; DB 4; Length 858;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 HDLHVH 8
Db      230 HELHVH 237

RESULT 3
US-09-255-829-29
; Sequence 29, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600

```

```

; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0130002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-255-829-29

Query Match      91.3%; Score 42; DB 4; Length 858;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 HDLHVH 8
Db      230 HELHVH 237

RESULT 4
US-09-255-829-20
; Sequence 20, Application US/09255829
; Patent No. 6461617
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/255,829
; FILING DATE: 23-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02273
; FILING DATE: 22-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,893
; FILING DATE: 27-DEC-1996

```

ATTORNEY/AGENT INFORMATION:  
NAME: ESMOND, ROBERT W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 1581.0130002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1169 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-255-829-20

Query Match 91.3%; Score 42; DB 4; Length 1169;  
Best Local Similarity 87.5%; Pred. No. 18;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLHVLH 8  
|:|||||  
Db 230 HELHVLH 237

RESULT 5  
US-08-913-880C-1  
Sequence 1, Application US/08913880C  
Patent No. 6372225  
GENERAL INFORMATION:  
APPLICANT: MATSUDA, Morihito  
TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS  
FILE REFERENCE: 216-380P  
CURRENT APPLICATION NUMBER: US/08/913,880C  
CURRENT FILING DATE: 1997-11-24  
NUMBER OF SEQ ID NOS: 17  
SEQ ID NO 1  
LENGTH: 1315  
TYPE: PRT  
ORGANISM: Clostridium tetani  
US-08-913-880C-1

Query Match 91.3%; Score 42; DB 4; Length 1315;  
Best Local Similarity 87.5%; Pred. No. 20;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLHVLH 8  
|:|||||  
Db 233 HELHVLH 240

RESULT 6  
US-09-198-452A-306  
Sequence 306, Application US/09198452A  
Patent No. 6559294  
GENERAL INFORMATION:  
APPLICANT: Griffiths, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
thereof and uses thereof, in particular for the diagnosis, prevention  
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
TITLE OF INVENTION: and treatment of infection  
FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198,452A  
CURRENT FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 306  
LENGTH: 907  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-306

Query Match 82.6%; Score 38; DB 4; Length 907;  
Best Local Similarity 62.5%; Pred. No. 65;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HDLHVLH 8  
|:|||||  
Db 541 HDLHVLH 548

RESULT 7  
US-08-118-270-257  
Sequence 257, Application US/08118270  
Patent No. 5508384  
GENERAL INFORMATION:  
APPLICANT: Murphy, Randall B.  
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
NUMBER OF SEQUENCES: 348  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/118,270  
FILING DATE: 09-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/943,236  
FILING DATE: 10-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, Kevin G.  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: MURPHY-2A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 257:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-118-270-257

Query Match 76.1%; Score 35; DB 1; Length 60;  
Best Local Similarity 62.5%; Pred. No. 16;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HDLHVLH 8  
|:|||||  
Db 42 HDLHVLH 49

RESULT 8  
PCT-US93-08528-257  
Sequence 257, Application PC/TUS9308528  
GENERAL INFORMATION:  
APPLICANT: New York University  
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
NUMBER OF SEQUENCES: 348  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.

```
/ COUNTRY: USA
/ ZIP: 20004
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US93/08528
/ FILING DATE: 09-SEP-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/943,236
/ FILING DATE: 10-SEP-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Townsend, Kevin G.
/ REGISTRATION NUMBER: 34,033
/ REFERENCE/DOCKET NUMBER: MURPHY=2 PCT
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-628-5197
/ TELEFAX: 202-737-3528
/ * TELEX: 248633
/ INFORMATION FOR SEQ ID NO: 257:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 60 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ PCT-US93-08528-257

Query Match 76.1%; Score 35; DB 5; Length 60;
Best Local Similarity 62.5%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 42 HDLVSILH 49

RESULT 9
PCT-US93-05000-31
/ Sequence 31, Application PC/TUS9305000
/ GENERAL INFORMATION:
/ APPLICANT: MITOTIX
/ TITLE OF INVENTION: D-Type Cyclin and Uses Related Thereto
/ NUMBER OF SEQUENCES: 42
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
/ STREET: Two Militia Drive
/ CITY: Lexington
/ STATE: Massachusetts
/ COUNTRY: US
/ ZIP: 02173
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US93/05000
/ FILING DATE: 19930525
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/07/888,178
/ FILING DATE: 26-MAY-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Granahan, Patricia
/ REGISTRATION NUMBER: 32,227
/ REFERENCE/DOCKET NUMBER: CSHL91-02A
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617-861-6240
/ TELEFAX: 616-861-9540
/ INFORMATION FOR SEQ ID NO: 31:

/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 269 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: unknown
/ MOLECULE TYPE: protein
/ PCT-US93-05000-31

Query Match 76.1%; Score 35; DB 5; Length 269;
Best Local Similarity 62.5%; Pred. No. 66;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 158 HDFIYLH 165

RESULT 10
US-09-107-532A-7103
/ Sequence 7103, Application US/09107532A
/ Patent No. 6583275
/ GENERAL INFORMATION:
/ APPLICANT: Lynn A Doucette-Stamm and David Bush
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
/ ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
/ NUMBER OF SEQUENCES: 7310
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: GENOME THERAPEUTICS CORPORATION
/ STREET: 100 Beaver Street
/ CITY: Waltham
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02354
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: CD-ROM ISO9660
/ COMPUTER: PC
/ OPERATING SYSTEM: <Unknown>
/ SOFTWARE: ASCII
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/107,532A
/ FILING DATE: 30-Jun-1998
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/085,598
/ FILING DATE: 14 May 1998
/ APPLICATION NUMBER: 60/051571
/ FILING DATE: July 2, 1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Arinello, Pamela Deneke
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-012
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-8277
/ INFORMATION FOR SEQ ID NO: 7103:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 204 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Enterococcus faecium
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATION 1...204
/ SEQUENCE DESCRIPTION: SEQ ID NO: 7103:
US-09-107-532A-7103

Query Match 73.9%; Score 34; DB 4; Length 204;
Best Local Similarity 75.0%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
| | | | |
```

Db 107 HGLYHVHL 114

RESULT 11  
US-08-117-083-20  
; Sequence 20, Application US/08117083  
; Patent No. 5719054  
; GENERAL INFORMATION:  
; APPLICANT: Boursnell, Michael E.  
; APPLICANT: Inglis, Stephen C.  
; APPLICANT: Munro, Alan J.  
; TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human  
; TITLE OF INVENTION: Papilloma Virus Proteins  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Walter H. Dreger  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/117,083  
; FILING DATE: 10-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-58783  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-781-1989  
; TELEFAX: 415-398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 60 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..60  
; OTHER INFORMATION: /note= "Xaa refers to stop codon in  
; OTHER INFORMATION: the open reading frame."  
US-08-117-083-20

Query Match 71.7%; Score 33; DB 1; Length 60;  
Best Local Similarity 85.7%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 DLHVLH 8  
Db 24 DLHVLH 30

RESULT 12  
US-07-879-685B-1  
; Sequence 1, Application US/07879685B  
; Patent No. 5296383  
; GENERAL INFORMATION:  
; APPLICANT: DAIKIN INDUSTRIES, LTD.  
; TITLE OF INVENTION: A human centromere antigen  
; TITLE OF INVENTION: polypeptide  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Umeda Center Building, 4-12  
; STREET: Nakazaki-nishi, 2-chome

; CITY: Kita-ku  
; STATE: Osaka  
; COUNTRY: Japan  
; ZIP: 530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/879,685B  
; FILING DATE: 19920507  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 3-102517  
; FILING DATE: 08-May-1991  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 65 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: Internal fragment  
; ORIGINAL SOURCE:  
; ORGANISM: human  
US-07-879-685B-1

Query Match 71.7%; Score 33; DB 1; Length 65;  
Best Local Similarity 83.3%; Pred. No. 37;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHV 6  
Db 42 HDLVEV 47

RESULT 13  
US-07-879-685B-4  
; Sequence 4, Application US/07879685B  
; Patent No. 5296383  
; GENERAL INFORMATION:  
; APPLICANT: DAIKIN INDUSTRIES, LTD.  
; TITLE OF INVENTION: A human centromere antigen  
; TITLE OF INVENTION: polypeptide  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Umeda Center Building, 4-12  
; STREET: Nakazaki-nishi, 2-chome  
; CITY: Kita-ku  
; STATE: Osaka  
; COUNTRY: Japan  
; ZIP: 530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/879,685B  
; FILING DATE: 19920507  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 3-102517  
; FILING DATE: 08-May-1991  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 162 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-879-685B-4

Query Match 71.7%; Score 33; DB 1; Length 162;



Best Local Similarity 83.3%; Pred. No. 87;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHV 6  
|||:|  
Db 139 HDLHV 144

## RESULT 14

US-08-117-083-62  
; Sequence 62, Application US/08117083  
; Patent No. 5719054  
; GENERAL INFORMATION:  
; APPLICANT: Bourns, Michael E.  
; APPLICANT: Inglis, Stephen C.  
; APPLICANT: Munro, Alan J.  
; TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human  
; TITLE OF INVENTION: Papilloma Virus Proteins  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Walter H. Dreger  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/117,083  
; FILING DATE: 10-SEP-1993  
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-58783  
; TELEPHONE: 415-781-1989  
; TELEFAX: 415-398-3249  
; TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 416 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..416  
; OTHER INFORMATION: /note="Xaa refers to stop codon in  
; the open reading frame."  
US-08-117-083-62

Query Match 71.7%; Score 33; DB 1; Length 416;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 DLIHVLH 8  
|||:|  
Db 307 DLIHFLH 313

## RESULT 15

US-08-311-023-2  
; Sequence 2, Application US/08311023  
; Patent No. 5693465  
; GENERAL INFORMATION:  
; APPLICANT: MANNING, David Lockwood  
; APPLICANT: NICHOLSON, Robert Ian

APPLICANT: GEE, Julia Margaret  
; APPLICANT: GREEN, Christopher Douglas  
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF  
; TITLE OF INVENTION: BREAST TUMOURS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Young & Thompson  
; STREET: 745 South 23rd Street  
; CITY: Arlington  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/311,023  
; FILING DATE: 22-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PATCH, Andrew J.  
; REGISTRATION NUMBER: Reg. No. 5693465 32,925  
; REFERENCE/DOCKET NUMBER: WCM.56  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703/521-2297  
; TELEFAX: 703/685-0573  
; TELEX: 248425  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 431 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-311-023-2

Query Match 71.7%; Score 33; DB 1; Length 431;  
Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8  
|||:|  
Db 232 HDYHILH 239

Search completed: November 12, 2003, 09:38:08  
Job time : 14 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:35:43 ; Search time 23 Seconds  
(without alignments)  
59.739 Million cell updates/sec

Title: US-10-064-903-2

Perfect score: 46

Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

```

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	100.0	8	15	US-10-064-903-2
2	42	91.3	439	14	US-10-011-588-17
3	42	91.3	441	11	US-09-910-346C-20
4	42	91.3	441	14	US-10-011-588-7
5	42	91.3	441	14	US-10-011-588-23
6	42	91.3	444	14	US-10-011-588-43
7	42	91.3	458	12	US-10-241-596-114
8	42	91.3	548	12	US-10-241-596-24
9	42	91.3	848	14	US-10-011-588-45
10	42	91.3	852	14	US-10-011-588-25
11	42	91.3	858	12	US-10-241-596-22
12	42	91.3	860	12	US-10-241-596-175
13	42	91.3	862	12	US-10-241-596-94
14	42	91.3	862	12	US-10-241-596-171
15	42	91.3	862	12	US-10-241-596-173

```

16 42 91.3 864 12 US-10-241-596-102 Sequence 102, App
17 42 91.3 865 12 US-10-241-596-100 Sequence 100, App
18 42 91.3 866 12 US-10-241-596-88 Sequence 88, App
19 42 91.3 866 12 US-10-241-596-104 Sequence 104, App
20 42 91.3 867 12 US-10-241-596-80 Sequence 80, App
21 42 91.3 867 12 US-10-241-596-96 Sequence 96, App
22 42 91.3 867 12 US-10-241-596-98 Sequence 98, App
23 42 91.3 870 12 US-10-241-596-92 Sequence 92, App
24 42 91.3 871 12 US-10-241-596-84 Sequence 84, App
25 42 91.3 871 12 US-10-241-596-86 Sequence 86, App
26 42 91.3 871 12 US-10-241-596-90 Sequence 90, App
27 42 91.3 872 12 US-10-241-596-145 Sequence 145, App
28 42 91.3 876 12 US-10-241-596-82 Sequence 82, App
29 42 91.3 876 12 US-10-241-596-106 Sequence 106, App
30 42 91.3 876 12 US-10-241-596-108 Sequence 108, App
31 42 91.3 879 12 US-10-241-596-143 Sequence 143, App
32 42 91.3 887 12 US-10-241-596-147 Sequence 147, App
33 42 91.3 888 12 US-10-241-596-112 Sequence 112, App
34 42 91.3 1169 12 US-10-241-596-20 Sequence 20, App
35 42 91.3 1315 12 US-10-241-596-141 Sequence 141, App
36 42 91.3 1420 12 US-10-241-596-110 Sequence 110, App
37 42 91.3 436 14 US-10-011-588-15 Sequence 15, App
38 38 82.6 443 14 US-10-011-588-39 Sequence 39, App
39 38 82.6 443 14 US-10-011-588-41 Sequence 41, App
40 36 78.3 422 14 US-10-011-588-13 Sequence 13, App
41 36 78.3 427 14 US-10-011-588-35 Sequence 35, App
42 36 78.3 804 14 US-10-011-588-37 Sequence 37, App
43 35 76.1 251 9 US-09-764-853-579 Sequence 579, App
44 35 76.1 302 12 US-10-259-165-72 Sequence 72, App
45 35 76.1 302 12 US-10-259-165-410 Sequence 410, App

```

## ALIGNMENTS

```

RESULT 1
US-10-064-903-2
; Sequence 2, Application US/10064903
; Publication No. US20030059912A1
; GENERAL INFORMATION:
; APPLICANT: Biotecon Gesellschaft für biotechnologische Entwicklung und Consulting
; APPLICANT: mbh
; TITLE OF INVENTION: HYBRID PROTEIN FOR INHIBITING THE DEGRANULATION OF MASTOCYTES ANI
; FILE REFERENCE: BIO-001PCT-CIP
; CURRENT APPLICATION NUMBER: US/10/064,903
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: US 09/700,540
; PRIOR FILING DATE: 2001-01-19
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Clostridium tetani
; US-10-064-903-2

```

```

Query Match 100.0%; Score 46; DB 15; Length 8;
Best Local Similarity 100.0%; Pred. No. 5.8e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 HDLIHVLH 8
Db 1 HDLIHVLH 8

```

```

RESULT 2
US-10-011-588-17
; Sequence 17, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody

```

```

; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic botulinum neurotoxin light chain of
; OTHER INFORMATION: serotype G based on wild-type Clostridium
; OTHER INFORMATION: botulinum sequence
US-10-011-588-17

```

```

Query Match          91.3%; Score 42; DB 14; Length 439;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 HDLIHVLH 8
   |:|||||
Db 229 HELIHLV 236

```

```

RESULT 3
US-09-910-346C-20
; Sequence 20, Application US/09910346C
; Publication No. US20030027752A1
; GENERAL INFORMATION:
; APPLICANT: STEWARD, LANCE E
; APPLICANT: FERNANDEZ-SALAS, ESTER
; APPLICANT: HERRINGTON, TODD M
; APPLICANT: AOKI, KEI R
; TITLE OF INVENTION: Leucine-based motif and clostridial neurotoxins
; FILE REFERENCE: D-2885CIP
; CURRENT APPLICATION NUMBER: US/09/910,346C
; CURRENT FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 09/620,840
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-09-910-346C-20

```

```

Query Match          91.3%; Score 42; DB 11; Length 441;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 HDLIHVLH 8
   |:|||||
Db 230 HELIHLV 237

```

```

RESULT 4
US-10-011-588-7
; Sequence 7, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard

```

```

; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic botulinum neurotoxin light chain of
; OTHER INFORMATION: serotype B based on wild-type Clostridium
; OTHER INFORMATION: botulinum sequence
US-10-011-588-7

```

```

Query Match          91.3%; Score 42; DB 14; Length 441;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 HDLIHVLH 8
   |:|||||
Db 230 HELIHLV 237

```

```

RESULT 5
US-10-011-588-23
; Sequence 23, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:22
US-10-011-588-23

```

```

Query Match          91.3%; Score 42; DB 14; Length 441;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 HDLIHVLH 8
   |:|||||

```

```
Db      230 HELIHVLH 237

RESULT 6
US-10-011-588-43
; Sequence 43, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:42
; NAME/KEY: UNSURE
; LOCATION: (442)...(443)
; OTHER INFORMATION: Any amino acid at each position
US-10-011-588-43

Query Match      91.3%; Score 42; DB 14; Length 444;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
       |:|||||
Db      230 HELIHVLH 237

RESULT 7
US-10-241-596-114
; Sequence 114, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 114
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-114

Query Match      91.3%; Score 42; DB 12; Length 548;
Best Local Similarity 87.5%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
       |:|||||
Db      230 HELIHVLH 237

RESULT 9
US-10-011-588-45
; Sequence 45, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
```

```
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 848
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:44
US-10-011-588-45

Query Match          91.3%; Score 42; DB 14; Length 848;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
       |:|||||
Db      230 HELIHVLH 237

RESULT 10
US-10-011-588-25
; Sequence 25, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; APPLICANT: Jensen, Melody
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR FILING DATE: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:24
US-10-011-588-25

Query Match          91.3%; Score 42; DB 14; Length 852;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
       |:|||||
Db      229 HELIHVLH 236

RESULT 11
US-10-241-596-22
; Sequence 22, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1993-02-23

; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Clostridium botulinum
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:24
US-10-011-588-25

Query Match          91.3%; Score 42; DB 14; Length 852;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
       |:|||||
Db      229 HELIHVLH 236

RESULT 12
US-10-241-596-175
; Sequence 175, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 175
; LENGTH: 860
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-241-596-175

Query Match          91.3%; Score 42; DB 12; Length 860;
Best Local Similarity 87.5%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVLH 8
       |:|||||
Db      230 HELIHVLH 237

RESULT 13
US-10-241-596-94
; Sequence 94, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
```

RESULT 14  
US-10-241-596-171  
; Sequence 171, Application US/10241596  
; Publication No. US20030166238A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbiological Research Authority  
; TITLE OF INVENTION: Recombinant Toxin Fragments  
; FILE REFERENCE: 1581.0130003  
; CURRENT APPLICATION NUMBER: US/10/241,596  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 09/255,829  
; PRIOR FILING DATE: 1999-02-23  
; PRIOR APPLICATION NUMBER: US 09/242,689  
; PRIOR FILING DATE: 1999-02-23  
; PRIOR APPLICATION NUMBER: PCT/GB97/02273  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 08/782,893  
; PRIOR FILING DATE: 1996-12-27  
; PRIOR APPLICATION NUMBER: GB 9625996.5  
; PRIOR FILING DATE: 1996-12-13  
; PRIOR APPLICATION NUMBER: GB 9617671.4  
; PRIOR FILING DATE: 1996-08-23  
; NUMBER OF SEQ ID NOS: 175  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 94  
; LENGTH: 862  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-10-241-596-94

Query Match 91.3%; Score 42; DB 12; Length 862;  
Best Local Similarity 87.5%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
; :|||||  
Db 232 HELIHVLH 239

RESULT 15  
US-10-241-596-173  
; Sequence 173, Application US/10241596  
; Publication No. US20030166238A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbiological Research Authority  
; TITLE OF INVENTION: Recombinant Toxin Fragments  
; FILE REFERENCE: 1581.0130003  
; CURRENT APPLICATION NUMBER: US/10/241,596  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 09/255,829  
; PRIOR FILING DATE: 1999-02-23  
; PRIOR APPLICATION NUMBER: US 09/242,689  
; PRIOR FILING DATE: 1999-02-23  
; PRIOR APPLICATION NUMBER: PCT/GB97/02273  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 08/782,893  
; PRIOR FILING DATE: 1996-12-27  
; PRIOR APPLICATION NUMBER: GB 9625996.5  
; PRIOR FILING DATE: 1996-12-13  
; PRIOR APPLICATION NUMBER: GB 9617671.4  
; PRIOR FILING DATE: 1996-08-23  
; NUMBER OF SEQ ID NOS: 175  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 171  
; LENGTH: 862  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-10-241-596-171

Query Match 91.3%; Score 42; DB 12; Length 862;  
Best Local Similarity 87.5%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
; :|||||  
Db 230 HELIHVLH 237

RESULT 15  
US-10-241-596-173  
; Sequence 173, Application US/10241596  
; Publication No. US20030166238A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbiological Research Authority  
; TITLE OF INVENTION: Recombinant Toxin Fragments  
; FILE REFERENCE: 1581.0130003  
; CURRENT APPLICATION NUMBER: US/10/241,596  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 09/255,829  
; PRIOR FILING DATE: 1999-02-23  
; PRIOR APPLICATION NUMBER: US 09/242,689  
; PRIOR FILING DATE: 1999-02-23  
; PRIOR APPLICATION NUMBER: PCT/GB97/02273  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 08/782,893  
; PRIOR FILING DATE: 1996-12-27  
; PRIOR APPLICATION NUMBER: GB 9625996.5  
; PRIOR FILING DATE: 1996-12-13  
; PRIOR APPLICATION NUMBER: GB 9617671.4  
; PRIOR FILING DATE: 1996-08-23  
; NUMBER OF SEQ ID NOS: 175  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 173  
; LENGTH: 862  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-10-241-596-173

Query Match 91.3%; Score 42; DB 12; Length 862;  
Best Local Similarity 87.5%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
; :|||||  
Db 230 HELIHVLH 237

Search completed: November 12, 2003, 09:44:49  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:33:02 ; Search time 13.5 Seconds  
(without alignments)  
56,989 Million cell updates/sec

Title: US-10-064-903-2  
Perfect score: 46  
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	91.3	1268	2 S33411	botulinum neurotox
2	42	91.3	1291	2 A48940	bontoxilysin (EC 3
3	42	91.3	1291	2 I40631	non-proteolytic bo
4	42	91.3	1297	2 S3791	neurotoxin - Clost
5	42	91.3	1315	1 BTCLTN	tentoxilysin (EC 3
6	38	82.6	925	2 A72096	CT234 hypothetical
7	38	82.6	925	2 E81573	conserved hypothet
8	38	82.6	925	2 E86527	CT234 hypothetical
9	38	82.6	1274	2 I40813	neurotoxin type F
10	36	78.3	232	2 C85585	unknown protein en
11	36	78.3	232	2 B90735	hypothetical prote
12	36	78.3	268	2 S73042	purine nucleoside
13	36	78.3	489	2 G86867	prophage ps3 prote
14	36	78.3	773	2 C84554	hypothetical prote
15	36	78.3	1251	2 JH0256	botulinum neurotox
16	36	78.3	1352	2 S21178	botulinum neurotox
17	35	76.1	312	2 C71806	hypothetical prote
18	35	76.1	312	2 G64712	toxR-activated gen
19	35	76.1	416	2 T45051	hypothetical prote
20	35	76.1	431	2 T18753	hypothetical prote
21	35	76.1	500	2 C75455	carboxypeptidase-r
22	35	76.1	679	2 H95036	glycosyl hydrolase
23	35	76.1	737	2 D79507	alpha-xylosidase (
24	34	73.9	198	2 F95194	recombination prot
25	34	73.9	198	2 C98061	recombination prot
26	34	73.9	198	2 AD1788	recombination prot
27	34	73.9	198	2 AE1412	recombination prot
28	34	73.9	199	2 JCS718	superoxide dismuta
29	34	73.9	421	2 C84555	hypothetical prote

ALIGNMENTS

RESULT 1

S33411 botulinum neurotoxin type F - Clostridium barati

C:Species: Clostridium barati

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Jul-1999  
C:Accession: S33411; S31860

R:Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T  
FEMS Microbiol. Lett. 108, 175-182, 1993

A:Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin  
A:Reference number: S33411; MUID:93252228; PMID:8486245

A:Accession: S33411

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1268 <THO>

A:Cross-references: EMBL:X68262; NID:G49138; PIDN:CAA48329.1; PID:G49139

C:Superfamily: tetanus toxin

C:Keywords: neurotoxin

Query Match 91.3%; Score 42; DB 2; Length 1268;

Best Local Similarity 87.5%; Pred. No. 6.2;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8

Db 219 HELIHVLH 226

RESULT 2

A48940

bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum

N:Alternate names: botulinum neurotoxin type B (BoNT/B)

C:Species: Clostridium botulinum

C>Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999

C:Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08574

R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, N.P.

Appl. Environ. Microbiol. 58, 2345-2354, 1992

A:Title: Molecular cloning of the Clostridium botulinum structural gene encoding the ty

A:Reference number: A48940; MUID:92384550; PMID:1514783

A:Accession: A48940

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1291 <WHE>

A:Cross-references: GB:M81186; NID:G144734; PIDN:AAA23211.1; PID:G144735

A:Experimental source: type B, Danish

A>Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBI:P:112081); this publica

R:Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993

A:Title: Gene probes for identification of the botulin neurotoxin gene and specific i

A:Reference number: S48103; MUID:94013372; PMID:8408542

A:Accession: S48105

A>Status: preliminary

A:Molecule type: DNA

Qy 1 HDLHVLH 8  
| : | | | | |  
230 HELHVLH 237  
Db

RESULT 4  
S39791  
neurotoxin - Clostridium botulinum  
C,Species: Clostridium botulinum  
C,Date: 07-Oct-1994 #sequence\_revision 01-Dec-1995 #text\_change 16-Jul-1999  
C,Accession: S39791

Qy 1 HDLHVLH 8  
| : | | | | |  
230 HELHVLH 237  
Db



Qy 1 HDLIHVLH 8  
|:|||||  
Db 230 HELIHLVLH 237

RESULT 5  
BCL1N  
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani  
N:Alternate names: tetanus neurotoxin  
C:Species: Clostridium tetani  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 03-Jun-2002  
C:Accession: A25689; A25757; A25194; B25194; A60759; S69348; S09364  
R:Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Waller, U.; Hudel, M.  
EMBO J. 5, 2495-2502, 1986  
A:Title: Tetanus toxin: Primary structure, expression in E. coli, and homology with botu  
A:Reference number: A25689; MUID:87053814; PMID:3536478  
A:Accession: A25689  
A:Molecule type: DNA  
A:Residues: 1-1315 <EIS>  
A:Cross-references: GB:X04436; NID:G40769; PIDN:CAA28033.1; PID:G40770  
R:Fairweather, N.F.; Lyness, V.A.  
Nucleic Acids Res. 14, 7809-7812, 1986  
A:Title: The complete nucleotide sequence of tetanus toxin.  
A:Reference number: A25757; MUID:87040747; PMID:3774547  
A:Accession: A25757  
A:Molecule type: DNA  
A:Residues: 1-1315 <FAI>  
A:Cross-references: GB:X06214; NID:G40773; PIDN:CAA29564.1; PID:G40774  
A:Experimental source: strain CN3911  
R:Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.  
J. Bacteriol. 165, 21-27, 1986  
A:Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in E  
A:Reference number: A25194; MUID:86085672; PMID:3510187  
A:Accession: A25194  
A:Molecule type: DNA  
A:Residues: 743-1315 <FA2>  
A:Cross-references: GB:M12739; NID:G144920; PIDN:AAA23282.1; PID:G144921  
A:Accession: B25194  
A:Molecule type: protein  
A:Residues: 865-894 <PA3>  
R:Matsuda, M.; Lei, D.L.; Sugimoro, N.; Ozutsumi, K.; Okabe, T.  
Infect. Immun. 57, 3588-3593, 1989  
A:Title: Isolation, purification, and characterization of fragment B, the NH-2-terminal  
A:Reference number: A60759; MUID:90035436; PMID:2478476  
A:Accession: A60759  
A:Molecule type: protein  
A:Residues: 461-475 <MAT>  
R:Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.  
J. Immunol. 142, 394-402, 1989  
A:Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.  
A:Reference number: J50098; MUID:89093918; PMID:2463305  
A:Contents: annotation; epitope region  
R:Schiaivo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.  
Nature 359, 832-835, 1992  
A:Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt  
A:Reference number: S27125; MUID:93063293; PMID:1331807  
A:Contents: annotation  
R:de Filippis, V.; Vangelista, L.; Schiaivo, G.; Tonello, F.; Montecucco, C.  
Eur. J. Biochem. 229, 61-69, 1995  
A:Title: Structural studies on the zinc-endoropeptidase light chain of tetanus neurotoxin.  
A:Reference number: S69348; MUID:95262688; PMID:7744050  
A:Accession: S69348  
A:Molecule type: protein  
A:Residues: 2-31 <DEF>  
C:Comment: The source of this protein was an extrachromosomal plasmid.  
C:Comment: The precursor is cleaved by endogenous proteinase activity to form light (fra  
dual chains are not toxic when separated). The amino end of the heavy chain (fragment B)  
C:Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to ganglio  
C:Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internalized  
presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of sy  
A:Function:  
A:Description: blocks neuroexcitotoxicity via hydrolysis of a Gln-Phe peptide bond in synapt  
C:Superfamily: tetanus toxin

C:Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc  
F:2-457/Product: tentoxilysin light chain (fragment A) #status predicted <TTL>  
F:461-1315/Product: tentoxilysin heavy chain (fragment B.C) #status experimental <TTH>  
F:461-864/Domain: channel forming (fragment B) #status predicted <TXB>  
F:865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC>  
F:233,237/Binding site: zinc (His) #status predicted  
F:234/Active site: Glu #status predicted

Query Match 91.3%; Score 42; DB 1; Length 1315;  
Best Local Similarity 87.5%; Pred. No. 6.4;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8  
|:|||||  
Db 233 HELIHLVLH 240

RESULT 6  
A72096  
ct234 hypothetical protein - Chlamydophila pneumoniae (strain CWL029)  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Jun-2000  
C:Accession: A72096  
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606; PMID:10192388  
A:Accession: A72096  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-925 <ARN>  
A:Cross-references: GB:AB001614; GB:AB001363; NID:G4376562; PIDN:AA018442.1; PID:G43765  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: CPn0293  
C:Superfamily: Chlamydia trachomatis hypothetical protein CT234

Query Match 82.6%; Score 38; DB 2; Length 925;  
Best Local Similarity 62.5%; Pred. No. 25;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8  
|:|||||  
Db 538 HDLIHLH 545

RESULT 7  
E81573  
conserved hypothetical protein CP0465 [imported] - Chlamydophila pneumoniae (strain AR3;  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 20-Jun-2000  
C:Accession: E81573  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255; PMID:10684935  
A:Accession: E81573  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-925 <REA>  
A:Cross-references: GB:AE002208; GB:AE002161; NID:G7189387; PIDN:AAF38302.1; PID:G71893;  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CP0465  
C:Superfamily: Chlamydia trachomatis hypothetical protein CT234

Query Match 82.6%; Score 38; DB 2; Length 925;  
Best Local Similarity 62.5%; Pred. No. 25;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8  
|:|||||

Db 538 HDLLHITH 545

## RESULT 8

E86527

Cr234 hypothetical protein [imported] - Chlamydomophila pneumoniae (strain J138)  
 C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
 C:Accession: E86527  
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A>Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:Reference number: A86491; MUID:20330349; PMID:10871362  
 A:Accession: E86527  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-925 <STO>  
 A:Cross-references: GB:BA000008; NID:g8978667; PIDN:BAA98503.1; GSPDB:GN00142  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: CPJ0293  
 C:Superfamily: Chlamydia trachomatis hypothetical protein Cr234

Query Match 82.6%; Score 38; DB 2; Length 925;  
 Best Local Similarity 62.5%; Pred. No. 25;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 |||||  
 Db 538 HDLLHITH 545

## RESULT 9

I40813  
 neurotoxin type F - Clostridium botulinum  
 C:Species: Clostridium botulinum  
 C>Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Jul-1999  
 R:East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E  
 FEMS Microbiol. Lett. 96, 225-230, 1992  
 A>Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.  
 A:Reference number: 140644  
 A:Accession: I40813  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1274 <RES>  
 A:Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867  
 R:Campbell, K.D.; Collins, M.D.; East, A.K.  
 J. Clin. Microbiol. 31, 2255-2262, 1993  
 A>Title: Gene probes for identification of the botulinum neurotoxin gene and specific id  
 A:Reference number: S48103; MUID:94013372; PMID:8408542  
 A:Accession: S48108  
 A:Status: preliminary; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 634-1002 <CAM>  
 A:Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789  
 C:Superfamily: tetanus toxin  
 C:Keywords: neurotoxin

Query Match 82.6%; Score 38; DB 2; Length 1274;  
 Best Local Similarity 75.0%; Pred. No. 36;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 |||||  
 Db 227 HELIHALH 234

## RESULT 10

C85585  
 unknown protein encoded by prophage CP-933K [imported] - Escherichia coli (strain O157:H  
 C:Species: Escherichia coli  
 C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: C85585  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca  
 Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: C85585  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-232 <STO>  
 A:Cross-references: GB:AE005174; NID:g12513758; PIDN:AAG55143.1; GSPDB:GN00145; UWGP:ZO  
 A:Experimental source: strain O157:H7, substrain EDL533  
 C:Genetics:  
 A:Gene: Z0990

Query Match 78.3%; Score 36; DB 2; Length 232;  
 Best Local Similarity 62.5%; Pred. No. 14;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 |||||  
 Db 142 HELHVFH 149

## RESULT 11

B90735  
 hypothetical protein ECs0850 [imported] - Escherichia coli (strain O157:H7, substrain F  
 C:Species: Escherichia coli  
 C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C:Accession: B90735  
 R:Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.C  
 Gasawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ger  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: B90735  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-232 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA834273.1; PID:g13360309; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECs0850

Query Match 78.3%; Score 36; DB 2; Length 232;  
 Best Local Similarity 62.5%; Pred. No. 14;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 |||||  
 Db 142 HELHVFH 149

## RESULT 12

S73042  
 purine nucleoside phosphorylase pnph - Mycobacterium leprae  
 N:Alternate names: L308\_F2\_56 protein  
 C:Species: Mycobacterium leprae  
 C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 23-Mar-2001  
 C:Accession: S73042  
 R:Smith, D.R.; Robinson, K.  
 submitted to the EMBL Data Library, November 1993  
 A:Description: Mycobacterium leprae cosmid L308.  
 A:Reference number: S72590  
 A:Accession: S73042  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-268 <SMI>  
 A:Cross-references: EMBL:U00022; NID:g467164; PIDN:AAA17341.1; PID:g467183  
 C:Genetics:  
 A:Gene: pnph  
 A:Start codon: GTG  
 C:Superfamily: purine-nucleoside phosphorylase

Query Match 78.3%; Score 36; DB 2; Length 268;  
 Best Local Similarity 75.0%; Pred. No. 16;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 ||| |||  
 Db 95 HDLRHVH 102

## RESULT 13

G86867  
 C:Species: Lactococcus lactis subsp. lactis (strain IL1403)  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: G86867  
 R:Botulin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich  
 Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss  
 A:Reference number: A86625; MUID:21235186; PMID:11337471  
 A:Accession: G86867  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-489 <STO>  
 A:Cross-references: GB:AR005176; PID:gn2724983; PIDN:AAK06041.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: ps311

Query Match 78.3%; Score 36; DB 2; Length 489;  
 Best Local Similarity 71.4%; Pred. No. 31;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 DLHVLH 8  
 ||| |||  
 Db 49 DLHVLH 55

## RESULT 14

C84554  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: C84554  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Ko, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon, L.;  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: C84554  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-773 <STO>  
 A:Cross-references: GB:AE002093; NID:G4926870; PIDN:AAD32950.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g17610  
 A:Map position: 2

Query Match 78.3%; Score 36; DB 2; Length 773;  
 Best Local Similarity 75.0%; Pred. No. 50;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 ||| |||  
 Db 10 HELIHSLL 17

## RESULT 15

JH0256  
 C:Species: Clostridium botulinum type E precursor - Clostridium butyricum  
 C:Date: 30-Jun-1992 #sequence\_revision 15-May-1998 #text\_change 16-Jul-1999

C:Accession: JH0256; S16145  
 R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.  
 Biochem. Biophys. Res. Commun. 183, 107-113, 1992  
 A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type  
 A:Reference number: JH0256; MUID:92181428; PMID:11543481  
 A:Accession: JH0256  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-27, 'B', 29-1251 <POU>  
 A:Cross-references: EMBL:X62088; NID:G40379  
 A:Experimental source: strains ATCC 43181 and ATCC 43755  
 R:Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N  
 J. Gen. Microbiol. 137, 519-525, 1991  
 A:Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E tox  
 A:Reference number: S16145; MUID:91237316; PMID:2033376  
 A:Accession: S16145  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-229, 'M', 231-252 <FUJ>  
 A:Cross-references: EMBL:X53180; NID:G40407; PIDN:CAA37321.1; PID:G40408  
 A:Experimental source: strain BL6340  
 C:Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release  
 C:Superfamily: tetanus toxin  
 C:Keywords: neurotoxin  
 F:2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>  
 F:423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>  
 F:412-426/Disulfide bonds: #status predicted

Query Match 78.3%; Score 36; DB 2; Length 1251;  
 Best Local Similarity 75.0%; Pred. No. 85;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8  
 ||| |||  
 Db 212 HELIHSLL 219

Search completed: November 12, 2003, 09:37:29  
 Job time : 14.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:31:57 ; Search time 10 Seconds  
(without alignments)  
37.621 Million cell updates/sec

Title: US-10-064-903-2  
Perfect score: 46  
Sequence: 1 HDLIHVLH 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	42	91.3	1290	1	BXB CLOBO
2	42	91.3	1296	1	BXG-CLOBO
3	42	91.3	1314	1	TEYX-CLOTE
4	38	82.6	1274	1	BXF_CLOBO
5	36	78.3	268	1	PUNA MYCLE
6	36	78.3	1250	1	BXE_CLOBO
7	36	78.3	1250	1	BXE_CLOBO
8	34	73.9	198	1	RECE_LISIN
9	34	73.9	198	1	RECE_LISIN
10	34	73.9	198	1	RECE_LISIN
11	34	73.9	199	1	SODC_HAEDU
12	34	73.9	446	1	EX7L_VIBCH
13	34	73.9	449	1	EX7L_SALTI
14	34	73.9	449	1	EX7L_SALTY
15	34	73.9	456	1	EX7L_ECO57
16	34	73.9	456	1	EX7L_ECOLI
17	34	73.9	458	1	TEGU_HCMVA
18	34	73.9	2241	1	TEGU_HCMVA
19	33	71.7	239	1	CENB_SHEEP
20	33	71.7	352	1	PRK1_SCHPO
21	33	71.7	459	1	EX7L_YERPE
22	33	71.7	541	1	UMES_YEAST
23	33	71.7	599	1	CENB_HUMAN
24	33	71.7	599	1	CENB_MOUSE
25	33	71.7	606	1	CENB_CRIGR
26	33	71.7	697	1	AD26_MOUSE
27	33	71.7	740	1	FAS_FNECA
28	33	71.7	942	1	AMPN_MANSE
29	32	69.6	91	1	YVBG_VACCC
30	32	69.6	239	1	MT04_ECOLI
31	32	69.6	477	1	GLGA_CLOAB
32	32	69.6	528	1	UGA3_YEAST
33	32	69.6	612	1	OCTC_RAT

RESULT 1					
BXB_CLOBO					
ID	BXB_CLOBO	STANDARD;	PRT;	1290	AA.
AC	P10844;	P10843;			
DT	01-JUL-1989	(Rel. 11, Created)			
DT	01-JUL-1993	(Rel. 26, Last sequence update)			
DT	15-SEP-2003	(Rel. 42, Last annotation update)			
DE	Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)				
DE	(Bontoxilysin B).				
GN	BOTB.				
OS	Clostridium botulinum.				
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;				
OC	Clostridium.				
OX	NCBI_TaxID=1491;				
OX	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92384550; PubMed=1514783;				
RA	Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,				
RA	Minton N.P.;				
RT	"Molecular cloning of the Clostridium botulinum structural gene				
RT	encoding the type B neurotoxin and determination of its entire				
RT	nucleotide sequence."				
RL	Appl. Environ. Microbiol. 58:2345-2354(1992).				
RN	[2]				
RP	SEQUENCE OF 35-245 FROM N.A.				
RC	STRAIN=NTC 7273;				
RA	Szabo E.A., Pemberton J.M., Desmarchelier P.M.;				
RL	Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 633-993 FROM N.A.				
RC	STRAIN=NTC 7273;				
RX	MEDLINE=94013372; PubMed=8408542;				
RA	Campbell K., East A.K., Collins M.D.;				
RT	"Gene probes for identification of the botulinum neurotoxin gene and				
RT	specific identification of neurotoxin types B, E, and F."				
RL	J. Clin. Microbiol. 31:2255-2262(1993).				
RN	[4]				
RP	SEQUENCE OF 1-44 AND 441-466.				
RC	STRAIN=657;				
RX	MEDLINE=89000987; PubMed=3139097;				
RA	Dasgupta B.R., Datta A.;				
RT	"Botulinum neurotoxin type B (strain 657): partial sequence and				
RT	similarity with tetanus toxin."				
RL	Biochimie 70:811-817(1988).				
RN	[5]				
RP	SEQUENCE OF 1-16 AND 441-458.				
RC	STRAIN=OKRA;				
RX	MEDLINE=85197963; PubMed=3888113;				
RA	Schmidt J.O., Sathyanarthy V., Dasgupta B.R.;				
RT	"Partial amino acid sequences of botulinum neurotoxins types B and				
RT	E."				
RL	Arch. Biochem. Biophys. 238:544-548(1985).				
RN	[6]				
RP	IDENTIFICATION AS ZINC-PROTEASE.				
RX	MEDLINE=93054694; PubMed=1429690;				
RA	Schiaivo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;				

P19821 thermus aqu  
P26607 escherichia  
P59342 shigella fl  
P53550 saccharomyc  
P21783 xenopus lae  
O66407 aquifex lae  
Q9VV96 drosophila  
Q8Y3A6 raietonia s  
Q9RKN3 streptomye  
Q98ET0 rhizobium l  
Q9HW5 thermoplasma  
Q9PIM9 chlamydia m

## ALIGNMENTS

"Botulinum neurotoxins are zinc proteins.";  
 J. Biol. Chem. 267:23479-23483(1992).  
 [7]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=93063293; PubMed=1331807;  
 RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,  
 RA Dasgupta B.R., Montecucco C.;  
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 by proteolytic cleavage of synaptobrevin.";  
 RL Nature 359:832-835(1992).  
 CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF  
 SYNAPTOSOMAL VESICLE.  
 CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 detected action on small molecule substrates.  
 CC -!- COPACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M81186; AAA23211.1; -;  
 DR EMBL; Z11934; CAA77991.1; -;  
 DR EMBL; X70817; CAA50148.1; -;  
 DR PIR; A48940; A48940.  
 DR PDB; 1EPW; 01-NOV-00.  
 DR PDB; 1F31; 01-NOV-00.  
 DR PDB; 1P82; 16-AUG-00.  
 DR PDB; 1P83; 16-AUG-00.  
 DR PDB; 1FQH; 06-DEC-00.  
 DR PDB; 1G9A; 13-NOV-02.  
 DR PDB; 1G9B; 13-NOV-02.  
 DR PDB; 1G9C; 13-NOV-02.  
 DR PDB; 1G9D; 13-NOV-02.  
 DR PDB; 111E; 21-NOV-01.  
 DR MEROPS; M27.002; -;  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR006025; Zn M27peptidase.  
 DR Pfam; PF01742; Peptidase M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;  
 3D-structure.  
 FT INIT MET 0  
 FT CHAIN 1 440 BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.  
 FT CHAIN 441 1290 BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.  
 FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 230 230 BY SIMILARITY.  
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 436 445 INTERCHAIN (PROBABILE).  
 FT DISULFID 29 29 T -> M (IN REF. 4).  
 FT CONFLICT 217 217 R -> G (IN REF. 2).  
 FT CONFLICT 224 224 A -> S (IN REF. 2).  
 FT CONFLICT 463 463 S -> R (IN REF. 4).  
 FT SEQUENCE 1290 AA; 150670 MW; D21746E2C024DF43 CRC64;  
 SQ  
 Query Match 91.3%; Score 42; DB 1; Length 1290;  
 Best Local Similarity 87.5%; Pred. No. 2.8;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HDLIHVLH 8  
 DB 229 HELIHLVH 236  
 :|:|:|:|:|:|  
 RESULT 2  
 EXG\_CLOBO STANDARD; PRT; 1296 AA.  
 AC Q60393;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)  
 DB Bontoxilysin G).  
 GN BONG.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=113 / 30;  
 RX MEDLINE=94092745; PubMed=8268233;  
 RA Campbell K., Collins M.D., East A.K.;  
 RT "Nucleotide sequence of the gene coding for Clostridium botulinum  
 (Clostridium argentinense) type G neurotoxin: genealogical comparison  
 with other clostridial neurotoxins".  
 RL Biochim. Biophys. Acta 1216:487-491(1993).  
 CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 ENDOPEPTIDASE.  
 CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 detected action on small molecule substrates.  
 CC -!- COPACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a  
 heavy chain (H). The light chain has the pharmacological activity,  
 while the N- and C-terminal of the heavy chain mediate channel  
 formation and toxin binding, respectively.  
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M81186; AAA23211.1; -;  
 DR EMBL; Z11934; CAA77991.1; -;  
 DR EMBL; X70817; CAA50148.1; -;  
 DR PIR; A48940; A48940.  
 DR PDB; 1EPW; 01-NOV-00.  
 DR PDB; 1F31; 01-NOV-00.  
 DR PDB; 1P82; 16-AUG-00.  
 DR PDB; 1P83; 16-AUG-00.  
 DR PDB; 1FQH; 06-DEC-00.  
 DR PDB; 1G9A; 13-NOV-02.  
 DR PDB; 1G9B; 13-NOV-02.  
 DR PDB; 1G9C; 13-NOV-02.  
 DR PDB; 1G9D; 13-NOV-02.  
 DR PDB; 111E; 21-NOV-01.  
 DR MEROPS; M27.002; -;  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR006025; Zn M27peptidase.  
 DR Pfam; PF01742; Peptidase M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;  
 3D-structure.  
 FT INIT MET 0  
 FT CHAIN 1 440 BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.  
 FT CHAIN 441 1290 BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.  
 FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 230 230 BY SIMILARITY.  
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 436 445 INTERCHAIN (PROBABILE).  
 FT DISULFID 29 29 T -> M (IN REF. 4).  
 FT CONFLICT 217 217 R -> G (IN REF. 2).  
 FT CONFLICT 224 224 A -> S (IN REF. 2).  
 FT CONFLICT 463 463 S -> R (IN REF. 4).  
 FT SEQUENCE 1290 AA; 150670 MW; D21746E2C024DF43 CRC64;  
 SQ

FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT SITE 230 230 BY SIMILARITY.  
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 435 449 INTERCHAIN (PROBABLE).  
 SQ SEQUENCE 1296 AA; 149013 MW; DC8547B15F665C31 CRC64;  
 Query Match 91.3%; Score 42; DB 1; Length 1296;  
 Best Local Similarity 87.5%; Pred. No. 2.8;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 HDLHVHL 8  
 Db 229 HELHVHL 236  
 RESULT 3  
 ID TCTX CLOTE STANDARD; PRT; 1314 AA.  
 AC P04958;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Tetanus toxin precursor (EC 3.4.24.68) (Pentoxylisin) [Contains:  
 DE Tetanus toxin light chain (Tetanus toxin chain L); Tetanus toxin heavy  
 DE chain (Tetanus toxin chain H)].  
 GN TCTX OR CTP60.  
 OS Clostridium tetani.  
 OG Plasmid pE88, and Plasmid 75 Kbp.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI TaxID=1513;  
 RN [1] \_SEQUENCE FROM N.A.  
 RC PLASMID=75 Kbp;  
 RX MEDLINE=87053814; PubMed=3536478;  
 RA Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,  
 RA Weller U., Hudel M., Habermann E., Niemann H.;  
 RT "Tetanus toxin: primary structure, expression in E. coli, and  
 RT homology with botulinum toxins.";  
 RL EMBO J. 5:2495-2502(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CN3911; PLASMID=75 Kbp;  
 RX MEDLINE=87040747; PubMed=3774547;  
 RA Fairweather N.F., Lyness V.A.;  
 RT "The complete nucleotide sequence of tetanus toxin.";  
 RL Nucleic Acids Res. 14:7809-7812(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Massachusetts / E88; PLASMID=pE88;  
 RX MEDLINE=22457253; PubMed=12552129;  
 RA Brueggemann H., Baumer S., Fricke W.F., Wierzer A., Liesegang H.,  
 RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,  
 RA Gottschalk G.;  
 RT "The genome sequence of Clostridium tetani, the causative agent of  
 RT tetanus disease.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).  
 RN [4]  
 RP SEQUENCE OF 742-1314 FROM N.A.  
 RC PLASMID=75 Kbp;  
 RX MEDLINE=86085672; PubMed=3510187;  
 RA Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;  
 RT "Cloning, nucleotide sequencing, and expression of tetanus toxin  
 RT fragment C in Escherichia coli.";  
 RL J. Bacteriol. 165:21-27(1986).  
 RN [5]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=90201034; PubMed=2108021;  
 RA Krieglstein K., Henschen A., Weller U., Habermann E.;  
 RT "Arrangement of disulfide bridges and positions of sulfhydryl groups  
 RT in tetanus toxin.";  
 RL Eur. J. Biochem. 188:39-45(1990).  
 RN [6]

RP PARTIAL SEQUENCE.  
 RX MEDLINE=92037649; PubMed=1935979;  
 RA Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;  
 RT "Limited proteolysis of tetanus toxin. Relation to activity and  
 RT identification of cleavage sites.";  
 RL Eur. J. Biochem. 202:41-51(1991).  
 RN [7]  
 RP IDENTIFICATION AS ZINC-PROTEASE.  
 RX MEDLINE=93010948; PubMed=1396558;  
 RA Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,  
 RA Montecucco C.;  
 RT "Tetanus toxin is a zinc protein and its inhibition of  
 RT neurotransmitter release and protease activity depend on zinc.";  
 RL EMBO J. 11:3577-3583(1992).  
 RN [8]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=93063293; PubMed=1331807;  
 RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,  
 RA Dasgupta B.R., Montecucco C.;  
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 RT by proteolytic cleavage of synaptobrevin.";  
 RL Nature 359:832-835(1992).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.  
 RX MEDLINE=97475217; PubMed=9334741;  
 RA Umland T.C., Wingert L.M., Swaminathan S., Furey W.F., Schmidt J.J.,  
 RA Sax M.;  
 RT "Structure of the receptor binding fragment HC of tetanus  
 RT neurotoxin.";  
 RL Nat. Struct. Biol. 4:788-792(1997).  
 CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77  
 CC BOND OF SYNAPTOSOMAL-2.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 76-Gln-|-Phe-77 bond in  
 CC synaptobrevin 2.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO  
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE  
 CC AND ARE NON-TOXIC AFTER SEPARATION.  
 CC -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO  
 CC GLYANGIOSIDE RECEPTORS.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X04436; CAA28033.1; -  
 CC EMBL; X06214; CAA29564.1; -  
 CC EMBL; AF528097; AAO37454.1; -  
 CC EMBL; M12739; AAA23282.1; -  
 CC EIR; A25689; BTCLTN.  
 CC PDB; 1AF9; 29-APR-98.  
 CC PDB; 1A8D; 14-OCT-98.  
 CC PDB; 1D0H; 27-MAR-00.  
 CC PDB; 1DFQ; 24-MAR-00.  
 CC PDB; 1DIW; 24-MAR-00.  
 CC PDB; 1DLL; 24-MAR-00.  
 CC PDB; 1FV3; 05-SEP-01.  
 CC MEROPS; M27.001; -  
 CC InterPro; IPR000395; Bontoxilysin.  
 CC InterPro; IPR006025; Zn\_MTPetase.  
 CC Pfam; PF01742; Peptidase\_M27; 1.  
 CC PRINTS; PR00760; BONTOXILYSIN.  
 CC ProDom; PD001963; Bontoxilysin; 1.

DR PROSITE; PS00142; ZINC PROTEASE; 1.  
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;  
FT 3D-structure; Complete proteome.

```
FT INIT_MET 0 0
FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
FT METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 233 233 BY SIMILARITY.
FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 669 689 INTERCHAIN.
FT DISULFID 438 466
FT DISULFID 1076 1092
FT HELIX 876 882
FT TURN 883 883
FT TURN 884 891
FT STRAND 884 891
FT TURN 892 893
FT STRAND 894 897
FT STRAND 904 907
FT TURN 909 910
FT STRAND 912 915
FT STRAND 920 925
FT TURN 928 929
FT STRAND 932 935
FT HELIX 938 940
FT TURN 941 946
FT STRAND 949 956
FT HELIX 962 968
FT TURN 969 970
FT STRAND 972 977
FT STRAND 980 981
FT HELIX 983 985
FT STRAND 987 995
FT TURN 996 997
FT STRAND 998 1004
FT TURN 1006 1007
FT STRAND 1010 1016
FT STRAND 1020 1020
FT TURN 1021 1022
FT STRAND 1031 1037
FT TURN 1039 1040
FT STRAND 1042 1047
FT TURN 1048 1049
FT STRAND 1050 1056
FT TURN 1058 1059
FT STRAND 1068 1074
FT TURN 1079 1080
FT STRAND 1082 1091
FT HELIX 1097 1105
FT TURN 1106 1107
FT STRAND 1112 1112
FT STRAND 1114 1114
FT TURN 1116 1117
FT STRAND 1120 1120
FT STRAND 1122 1122
FT TURN 1123 1124
FT STRAND 1127 1131
FT HELIX 1132 1134
FT TURN 1135 1136
FT STRAND 1137 1141
FT TURN 1144 1145
FT STRAND 1148 1152
FT STRAND 1155 1158
FT TURN 1159 1162
FT STRAND 1163 1166
FT STRAND 1173 1178
FT TURN 1184 1185
FT STRAND 1188 1188
FT STRAND 1190 1190
FT TURN 1191 1192
FT STRAND 1193 1201
```

Query Match

91.3%; Score 42; DB 1; Length 1314;

Best Local Similarity 87.5%; Pred No. 2.8; Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLHVLIH 8  
Db 232 HELIHLVH 239

## RESULT 4

ID\_BXF\_CLOBO STANDARD; PRT; 1274 AA.

AC B30396;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)

DE (Bontoxilysin F).

GN BOTF.

OS Clostridium botulinum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI\_TaxID=1491;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 23387;

RX MEDLINE=93012902; PubMed=1398040;

RA East A.K., Richardson P.T., Allaway D., Collins M.D.,

RA Roberts T.A., Thompson D.E.;

RT "Sequence of the gene encoding type F neurotoxin of Clostridium

RT botulinum.";

RL FEMS Microbiol. Lett. 75:225-230(1992).

RN [2]

RP SEQUENCE OF 1-64 FROM N.A.

RC STRAIN=Hobbs FT10;

RX MEDLINE=94297488; PubMed=7764998;

RA East A.K., Collins M.D.;

RT "Conserved structure of genes encoding components of botulinum

RT neurotoxin complex M and the sequence of the gene coding for the

RT nontoxic component in nonproteolytic Clostridium botulinum type F.";

RL Curr. Microbiol. 29:69-77(1994).

RN [3]

RP SEQUENCE OF 634-1002 FROM N.A.

RC MEDLINE=94013372; PubMed=8408542;

RA Campbell K., East A.K., Collins M.D.;

RT "Gene probes for identification of the botulin neurotoxin gene and

RT specific identification of neurotoxin types B, E, and F.";

RL J. Clin. Microbiol. 31:2255-2262(1993).

RN [4]

RP IDENTIFICATION OF SUBSTRATE.

RX MEDLINE=94230352; PubMed=8175689;

RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,

RA Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;

RT "Cleavage of members of the synaptobrevin/VAMP family by types D and

RT F botulin neurotoxins and tetanus toxin.";

RL J. Biol. Chem. 269:12764-12772(1994).

CC [1]

CC FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER

CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED

CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD

CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT

CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC

CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-LYS-59

CC BOND OF SYNAPTOSOMES-1 AND -2.

CC CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the

CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No

CC detected action on small molecule substrates.

CC COFACTOR: Binds 1 zinc ion per subunit (By similarity).

CC SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A

CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,

CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL

CC FORMATION AND TOXIN BINDING, RESPECTIVELY.

CC SUBCELLULAR LOCATION: Secreted.

CC MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF

CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

```

CC      -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M92906; AAA23283.1; -
DR      EMBL; S73676; AAC60475.1; -
DR      EMBL; X70820; CAA50151.1; -
DR      EMBL; X70816; CAA50147.1; -
DR      PIR; I40813; I40813.
DR      PIR; S48109; S48109.
DR      HSSP; F10845; 3BTA.
DR      MEROPS; M27.002; -.
DR      InterPro; IPR000395; Bontoxilysin.
DR      InterPro; IPR006025; Zn.MTpeptidase.
DR      Pfam; PF01742; Peptidase_M27; 1.
DR      ProDom; PD001963; Bontoxilysin; 1.
DR      PROSITE; PS00142; ZINC_PROTEASE; 1.
KW      Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT      CHAIN 1 436 BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT      CHAIN 2 227 BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT      METAL 227 227 ZINC (CATALYTIC) (BY SIMILARITY).
FT      ACT_SITE 228 228 BY SIMILARITY.
FT      METAL 231 231 ZINC (CATALYTIC) (BY SIMILARITY).
FT      DISULFID 429 445 INTERCHAIN (PROBABLE).
SQ      SEQUENCE 1274 AA; 146709 MW; 5899756A7438B921 CRC64;

Query Match      82.6%; Score 38; DB 1; Length 1274;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 HDLHVLR 8
DB      227 HELIHALH 234

RESULT 5
PUNA_MYCLE STANDARD; PRT; 268 AA.
AC P46862;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase)
DE (PNP).
GN PUNA OR DEOD OR ML0707 OR L308_F2_56.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltriwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."

```

---

```

RL      Nature 409:1007-1011 (2001).
CC      -!- FUNCTION: CLEAVAGE OF GUANOSINE OR INOSINE TO RESPECTIVE BASES AND
CC      SUGAR-1-PHOSPHATE MOLECULES (BY SIMILARITY).
CC      -!- CATALYTIC ACTIVITY: Purine nucleoside + phosphate = purine +
CC      alpha-D-ribose 1-phosphate.
CC      -!- PATHWAY: Purine nucleoside salvage.
CC      -!- SIMILARITY: BELONGS TO THE PNP/MTAP FAMILY 2 OF PHOSPHORYLASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U00022; AAA17341.1; -
DR      EMBL; AL583919; CAC30216.1; -
DR      PIR; S73042; S73042.
DR      HSSP; P81989; IQES.
DR      Leproma; ML0707; -.
DR      InterPro; IPR001369; Mtap_PNP.
DR      Pfam; PF00896; Mtap_PNP; 1.
DR      PROSITE; PS01240; PNP_MTAP_2; 1.
KW      Transferrase; Glycosyltransferase; Complete proteome.
SQ      SEQUENCE 268 AA; 27980 MW; 46C622532FC96A0F CRC64;

Query Match      78.3%; Score 36; DB 1; Length 268;
Best Local Similarity 75.0%; Pred. No. 7.2;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 HDLHVLR 8
DB      95 HDLRHVVR 102

RESULT 6
BXE_CLOBO STANDARD; PRT; 1250 AA.
AC Q00496;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE (Bontoxilysin E).
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Beluga;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulin neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92174922; PubMed=1541280;
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type-E
RT neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT gene.";
RL Eur. J. Biochem. 204:657-667(1992).
RN [3]
RP SEQUENCE OF 1-251 FROM N.A.
RX MEDLINE=90264400; PubMed=2160960;
RA Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. Biol. Chem. 265:9153-9158(1990).

```



[4]
RN SEQUENCE OF 1-13.
RP MEDLINE=85197963; PubMed=3888113;
RX Schmidt J.J., Sathiyamoorthy V., Dasgupta B.R.;
RA "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL Arch. Biochem. Biophys. 238:544-548(1985).
[5]
RN SEQUENCE OF 419-426.
RP MEDLINE=90344919; PubMed=2116911;
RX Gimenez J.A., Dasgupta B.R.;
RA "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT reveals the site trypsin nicks and homology with tetanus
RL neurotoxin.";
RL Biochimie 72:213-217(1990).
[6]
RN IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94063091; PubMed=8243676;
RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RX Benfenati F., Wilson M.C., Montecucco C.;
RA "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT COOH-terminal peptide bonds.";
RL FEBS Lett. 335:99-103(1993).
[7]
RN IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94124495; PubMed=8294407;
RA Binz T., Blas J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA Jahn R., Niemann H.;
RA "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";
RL J. Biol. Chem. 269:1617-1620(1994).
-!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
181 BOND IN SNAP-25.
-!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
detected action on small molecule substrates.
-!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
-!- SUBUNIT: DISULFIDE-LINKED HETEROIMER OF A LIGHT CHAIN (L) AND A
HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
FORMATION AND TOXIN BINDING, RESPECTIVELY.
-!- SUBCELLULAR LOCATION: Secreted.
-!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).
-----
EMBL; X62089; CAA43999.1; -
DR EMBL; X62683; CAA44558.1; -
DR PIR; S21178; S21178.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn\_MTPeptidase.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.

FT METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT\_SITE 212 212 BY SIMILARITY.
FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 411 425 INTERCHAIN (PROBABLE).
FT CONFLICT 176 176 R -> G (IN REF. 2).
FT CONFLICT 197 197 C -> S (IN REF. 2 AND 3).
FT CONFLICT 339 339 R -> A (IN REF. 2).
FT CONFLICT 772 772 I -> L (IN REF. 2).
FT CONFLICT 962 963 FE -> LQ (IN REF. 2).
FT CONFLICT 966 966 R -> A (IN REF. 2).
FT CONFLICT 1194 1194 N -> NN (IN REF. 2).
SQ SEQUENCE 1250 AA; 143712 MW; D9FCB26DDA041EB4 CRC64;
Query Match 78.3%; Score 36; DB 1; Length 1250;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Caps 0;
QY 1 HDLIHVLH 8
Db 211 HELIHSLS 218
RESULT 7
BXE\_CLOBU STANDARD; PRT; 1250 AA.
AC P30995; 1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 41, Last annotation update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BONT/E)
DE (Bontoxilysin E).
OS Clostridium butyricum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI\_TaxID=1492;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 43181, and ATCC 43755;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinum neurotoxin E derived from Clostridium
RL botulinum type E (strain Beluga) and Clostridium butyricum (strains
RL ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
[2]
RN SEQUENCE OF 1-251 FROM N.A.
RC STRAIN=BL6340;
RX MEDLINE=91237316; PubMed=2033376;
RA Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA Yokosawa N., Yashiki T., Oguma K.;
RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RL type E toxin gene from Clostridium butyricum strain BL6340.";
RL J. Gen. Microbiol. 137:519-525(1991).
[3]
RN SEQUENCE OF 1-48.
RC STRAIN=5262;
RA Gimenez J., Foley J., Dasgupta B.R.;
RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RL partial sequence and comparison.";
RL FASEB J. 2:A1750-A1750(1988).
-!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
ENDOPEPTIDASE.
-!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No
detected action on small molecule substrates.
-!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
-!- SUBUNIT: DISULFIDE-LINKED HETEROIMER OF A LIGHT CHAIN (L) AND A
HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL

Science 294:849-852(2001).

-!- FUNCTION: May play a role in DNA repair. It seems to be involved in an reBC-independent recombinational process of DNA repair. It may act with recF and recO (By similarity).

-!- SIMILARITY: BELONGS TO THE RECQ FAMILY.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL; AL596173; CAC98076.1; -

CC PIR; AD1788; AD1788.

CC ListList; LIN02850; -

CC HAMAP; MF\_00017; -; 1.

CC InterPro; IPR003583; HH1.

CC InterPro; IPR000093; RecR.

CC InterPro; IPR006171; Toprim\_dom.

CC InterPro; IPR006154; Toprim\_sub.

CC Pfam; PF02132; RecR; 1.

CC Pfam; PF01751; Toprim; 1.

CC SMART; SM00278; HhH1; 1.

CC SMART; SM00493; TOPRIM; 1.

CC TIGRFAMs; TIGR00615; recR; 1.

CC PROSITE; PS01300; RecR; 1.

CC DNA repair; DNA recombination; Zinc-finger; Complete proteome.

CC ZN FING 57 72 C4-TYPE (POTENTIAL).

CC SEQUENCE 198 AA; 21996 MW; 8A82E1A16415DFEF CRC64;

Query Match 73.9%; Score 34; DB 1; Length 198;

Best Local Similarity 75.0%; Pred. No. 12;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8

Db 101 HGLYHVLH 108

-----

RESULT 9

RECR\_LISMO STANDARD; PRT; 198 AA.

AC Q8Y3X7; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Recombination protein recR.

GN RECR OR LMO2702.

OS Listeria monocytogenes.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OC NCBI\_TaxID=1639;

EN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EGD-e / Serovar 1/2a;

RX MEDLINE=21537279; PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A., Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T., Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P., Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O., Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P., Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D., Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G., Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H., Nordisiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R., Remmel B., Rose M., Schluter T., Simoes N., Tierrez A., Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.; "Comparative genomics of Listeria species."

RT Science 294:849-852(2001).

CC -!- FUNCTION: May play a role in DNA repair. It seems to be involved in an reBC-independent recombinational process of DNA repair. It may act with recF and recO (By similarity).

CC

FORMATION AND TOXIN BINDING, RESPECTIVELY.

-!- SUBCELLULAR LOCATION: Secreted.

-!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL; X62088; CAA3398.1; -

CC EMBL; X53180; CAA37321.1; -

CC PIR; JH0256; JH0256.

CC HSSP; P10845; 3BTA.

CC MEROPS; M27.002; -

CC InterPro; IPR000395; Bontoxilysin.

CC InterPro; IPR006025; Zn Mtpetptase.

CC Pfam; PF01742; Peptidase M27; 1.

CC PRINTS; PR00760; BONTOXILYSIN.

CC ProDom; PD001963; Bontoxilysin; 1.

CC PROSITE; PS00142; ZINC PROTEINASE; 1.

CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.

CC INIT MET 0

CC CHAIN 0 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.

CC CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.

CC METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).

CC ACT\_SITE 212 212 BY SIMILARITY.

CC METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).

CC DISULFID 411 425 INTERCHAIN (PROBABLE).

CC CONFLICT 229 229 K -> M (IN REF. 2).

CC SEQUENCE 1250 AA; 143265 MW; 8171B5B2C312857 CRC64;

Query Match 78.3%; Score 36; DB 1; Length 1250;

Best Local Similarity 75.0%; Pred. No. 37;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8

Db 211 HGLHSLH 218

-----

RESULT 8

RECR\_LISIN STANDARD; PRT; 198 AA.

AC Q927D9; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Recombination protein recR.

GN RECR OR LIN2850.

OS Listeria innocua.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OC NCBI\_TaxID=1642;

EN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CLIP 11262 / Serovar 6a;

EX MEDLINE=21537279; PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A., Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T., Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P., Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O., Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P., Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D., Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G., Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H., Nordisiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R., Remmel B., Rose M., Schluter T., Simoes N., Tierrez A., Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.; "Comparative genomics of Listeria species."

RT

```

CC      -1- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AL591984; CAD00915.1; -
CC      PIR; AE1412; AE1412.
CC      Listlist; LMO02702; -.
CC      HAMAP; MF 00017; -.
CC      InterPro; IPR003583; HHH 1.
CC      InterPro; IPR000093; RecR.
CC      InterPro; IPR006171; Toprim dom.
CC      InterPro; IPR006154; Toprim_sub.
CC      Pfam; PF02132; RecR; 1.
CC      Pfam; PF01751; Toprim; 1.
CC      SMART; SM00278; HHH1; 1.
CC      SMART; SM00493; TOPRIM; 1.
CC      TIGRFAMs; TIGR00615; recR; 1.
CC      PROSITE; PS01300; RECR; 1.
CC      DNA repair; Zinc-finger; Complete proteome.
CC      ZN FING 57 72 C4-TYPE (POTENTIAL).
CC      SEQUENCE 198 AA; 21934 MW; E542E27BC3D05036 CRC64;
CC      Query Match 73.9%; Score 34; DB 1; Length 198;
CC      Best Local Similarity 75.0%; Pred. No. 12;
CC      Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CC      QY 1 HDLIHVLH 8
CC      Db 101 HGLYHVLH 108
CC      RESULT 10
CC      RECR STRPN STANDARD; PRT; 198 AA.
CC      AC Q9ZHC4;
CC      DT 30-MAY-2000 (Rel. 39, Created)
CC      DT 30-MAY-2000 (Rel. 39, Last sequence update)
CC      DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC      DE Recombination protein recR.
CC      GN RECR OR RECM OR SPI672 OR SPI672 OR SPI672.
CC      OS Streptococcus pneumoniae, and
CC      OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
CC      OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC      OC Streptococcus.
CC      OC NCBI_TaxID=1313, 171101;
CC      RN [1]
CC      SEQUENCE FROM N.A.
CC      RC STRAIN=G54 / Type 19F;
CC      RX MEDLINE=99061199; PubMed=9846742;
CC      RA Massida O., Anderluzzi D., Friedli L., Peger G.;
CC      RT "Unconventional organization of the division and cell wall gene
CC      cluster of Streptococcus pneumoniae.";
CC      RL Microbiology 144:3069-3078(1998).
CC      RN [2]
CC      SEQUENCE FROM N.A.
CC      RC STRAIN=339 and PN94-661;
CC      RX MEDLINE=20073037; PubMed=10605111;
CC      RA Enright M.C., Spratt B.G.;
CC      RT "Extensive variation in the ddl gene of penicillin-resistant
CC      Streptococcus pneumoniae results from a hitchhiking effect driven by
CC      the penicillin-binding protein 2b gene.";
CC      RL Mol. Biol. Evol. 16:1687-1695(1999).
CC      RN [3]
CC      SEQUENCE FROM N.A.
CC      RC STRAIN=ATCC BAA-334 / TIGR4;
CC      RX MEDLINE=21357209; PubMed=11463916;
CC      RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,

```

```

RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., C.L.,
RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen D., C.L.,
RA McDonald L.E., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Lofcus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RA "Complete genome sequence of a virulent isolate of Streptococcus
RA pneumoniae.";
RA Science 293:498-506(2001).
RA [4]
RA SEQUENCE FROM N.A.
RA STRAIN=ATCC BAA-255 / R6;
RA MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RA "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RA J. Bacteriol. 183:5709-5717(2001).
RA -1- FUNCTION: MAY PLAY A ROLE IN DNA REPAIR. IT SEEMS TO BE INVOLVED
CC IN AN REBC-DEPENDENT RECOMBINATIONAL PROCESS OF DNA REPAIR. IT
CC MAY ACT WITH RECF AND RECO (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RECR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AF068901; AAC95434.1; -
CC      EMBL; AJ243056; CAB64474.1; -
CC      EMBL; AJ243057; CAB64478.1; -
CC      EMBL; AE007460; AAK75751.1; -
CC      EMBL; AE008520; AAL00320.1; -
CC      PIR; C98061; C98061.
CC      PIR; F95194; F95194.
CC      TIGR; SPI672; -.
CC      HAMAP; MF 00017; -.
CC      InterPro; IPR000093; RecR.
CC      InterPro; IPR006171; Toprim dom.
CC      InterPro; IPR006154; Toprim_sub.
CC      Pfam; PF02132; RecR; 1.
CC      Pfam; PF01751; Toprim; 1.
CC      SMART; SM00493; TOPRIM; 1.
CC      TIGRFAMs; TIGR00615; recR; 1.
CC      PROSITE; PS01300; RECR; 1.
CC      DNA repair; DNA recombination; Zinc-finger; Complete proteome.
CC      ZN FING 57 72 C4-TYPE (POTENTIAL).
CC      SEQUENCE 198 AA; 21689 MW; FC6F0E98E333752 CRC64;
CC      Query Match 73.9%; Score 34; DB 1; Length 198;
CC      Best Local Similarity 75.0%; Pred. No. 12;
CC      Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CC      QY 1 HDLIHVLH 8
CC      Db 101 HGLYHVLH 108
CC      RESULT 11
CC      SODC-HAEDU STANDARD; PRT; 199 AA.
CC      ID SODC-HAEDU STANDARD; PRT; 199 AA.
CC      AC Q59452; Q59449; Q59453;
CC      DT 15-DEC-1998 (Rel. 37, Created)

```

15-DEC-1998 (Rel. 37, Last sequence update)  
15-SEP-2003 (Rel. 42, Last annotation update)  
DE Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1).  
GN SODC.  
OS Haemophilus ducreyi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=730;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=35000;  
RX MEDLINE=97149276; PubMed=8996084;  
RA Langford P.R., Kroll J.S.;  
RT "Distribution, cloning, characterisation and mutagenesis of sodC, the  
RT gene encoding copper/zinc superoxide dismutase, a potential  
RT determinant of virulence, in Haemophilus ducreyi.";  
RL FEMS Immunol. Med. Microbiol. 17:235-242(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=35000;  
RX MEDLINE=97149276; PubMed=8996084;  
RA Stevens M.K., Hassett D.J., Radolf J.D., Hansen E.J.;  
RT "Cloning and sequencing of the gene encoding the Cu,Zn-superoxide  
RT dismutase of Haemophilus ducreyi.";  
RL Gene 183:35-40(1996).  
RN [3]  
RP SEQUENCE OF 100-186 FROM N.A.  
RC STRAIN=35000;  
RX MEDLINE=96118708; PubMed=7496539;  
RA Kroll J.S., Langford P.R., Wilks K.E., Keil A.D.;  
RT "Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct  
RT from the eukaryotic enzyme, and not so rare after all";  
RL Microbiology 141:2271-2279(1995).  
CC -!- FUNCTION: Destroys radicals which are normally produced within the  
CC cells and which are toxic to biological systems. May play a role  
CC in the interactive biology of organisms with their hosts and so  
CC contribute to their capacity to cause disease.  
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
CC -!- COFACTOR: Binds 1 copper ion and 1 zinc ion per subunit (By  
CC similarity).  
CC -!- SUBUNIT: Homodimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Periplasmic.  
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
CC  
CC -----  
CC EMBL; X98737; CAA67289.1; --  
CC DR EMBL; U47664; BAB41293.1; --  
CC DR EMBL; X83125; CAA58206.1; --  
CC DR PIR; JC5718; JCS718.  
CC DR HSP; P24702; 2APS.  
CC DR InterPro; IPR001424; SOD\_CU\_ZN.  
CC DR Pfam; PF00080; sodcu; 1; \_CN\_ZN.  
CC DR ProDom; PD000469; SOD\_CU\_ZN; 1.  
CC DR PROSITE; PS00087; SOD\_CU\_ZN\_1; 1.  
CC DR PROSITE; PS00332; SOD\_CU\_ZN\_2; 1.  
CC DR Antioxidant; Oxidoreductase; Metal-binding; Copper; Zinc; Periplasmic;  
CC signal.  
CC  
CC FT FT SIGNAL 1 22 POTENTIAL.  
CC FT CHAIN 23 199 SUPEROXIDE DISMUTASE [CU-ZN].  
CC FT METAL 92 92 COPPER (BY SIMILARITY).  
CC FT METAL 94 94 COPPER (BY SIMILARITY).  
CC FT METAL 117 117 COPPER AND ZINC (BY SIMILARITY).  
CC FT METAL 126 126 ZINC (BY SIMILARITY).  
CC FT METAL 135 135 ZINC (BY SIMILARITY).  
CC FT METAL 138 138 ZINC (BY SIMILARITY).  
CC FT METAL 173 173 COPPER (BY SIMILARITY).  
CC

```

Query Match      73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVL 7
DB      154 HDLIHVL 160

RESULT 13
EX7L_SALTY
ID EX7L_SALTY STANDARD; PRT; 449 AA.
AC Q824Q1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR STV2753 OR T0345.
OS Salmonella typhi
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churche C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyonni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AL627275; CAB02714.1; -.
CC HAMAP; MF 00378; -.
CC InterPro; IPR003753; Exonuc VII_L.
CC Pfam; PF02601; Exonuc VII_L; 1.
CC TIGRFAMs; TIGR00237; xsea; 1.
CC Hydroxylase; Nuclease; Exonuclease; Complete proteome.
CC SEQUENCE 449 AA; 50720 MW; 511957D8C878F5D2 CRC64;

Query Match      73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVL 7
DB      154 HDLIHVL 160

RESULT 14
EX7L_SALTY
ID EX7L_SALTY STANDARD; PRT; 449 AA.
AC Q82N58;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR STM2512.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE008813; AAL21406.1; -.
CC HAMAP; MF 00378; -.
CC InterPro; IPR003753; Exonuc VII_L.
CC Pfam; PF02601; Exonuc VII_L; 1.
CC TIGRFAMs; TIGR00237; xsea; 1.
CC Hydroxylase; Nuclease; Exonuclease; Complete proteome.
CC SEQUENCE 449 AA; 50613 MW; 85356CE8560E161E CRC64;

Query Match      73.9%; Score 34; DB 1; Length 449;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 HDLIHVL 7
DB      154 HDLIHVL 160

```

```
RESULT 15
EX7L_ECO57
ID EX7L_ECO57 STANDARD; PRT; 456 AA.
AC OXABO;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable exodeoxyribonuclease VII large subunit (EC 3.1.11.6)
DE (Exonuclease VII large subunit).
GN XSEA OR Z3773 OR EGS371.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=83334;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Borna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Qnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5' to 3'
CC or 3' to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005480; AAG57620.1; -.
DR EMBL; AP002561; BAB36794.1; -.
DR FIC; C91050; C91050.
DR HAMAP; MF_00378; -.
DR InterPro; IPR003753; Exonuc_VII_L.
DR InterPro; IPR004365; tRNA_anti.
DR Pfam; PF02601; Exonuc_VII_L; 1.
DR Pfam; PF01336; tRNA_anti; 1.
DR TIGRFAMs; TIGR00237; xsea; 1.
DR Hydrolase; Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 456 AA; 51734 MW; 174EAE7F2EB3C37 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 456;
Best Local Similarity 71.4%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 HDLHVL 7
||:||||
```

Db 154 HDLHVL 160

Search completed: November 12, 2003, 09:35:40  
Job time : 11 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2003, 09:32:27 ; Search time 28 Seconds  
(without alignments)  
73,729 Million cell updates/sec

Title: US-10-064-903-2

Perfect score: 46

Sequence: 1 HDLHVH 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23.\*  
1: sp archaea.\*  
2: sp bacteria.\*  
3: sp fungi.\*  
4: sp human.\*  
5: sp invertebrate.\*  
6: sp mammal.\*  
7: sp mhc.\*  
8: sp organelle.\*  
9: sp phage.\*  
10: sp plant.\*  
11: sp rodent.\*  
12: sp virus.\*  
13: sp vertebrate.\*  
14: sp unclassified.\*  
15: sp virus.\*  
16: sp bacteriophage.\*  
17: sp archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	91.3	451	Q9R631	Q9R631 clostridium
2	42	91.3	1268	Q45851	Q45851 clostridium
3	42	91.3	1291	Q9ZAJ8	Q9ZAJ8 clostridium
4	42	91.3	1291	Q9JG71	Q9JG71 clostridium
5	42	91.3	1291	Q9J3K0	Q9J3K0 clostridium
6	42	91.3	1291	Q9JG77	Q9JG77 clostridium
7	42	91.3	1291	Q9GR96	Q9GR96 clostridium
8	42	91.3	1310	Q9JN27	Q9JN27 clostridium
9	39	84.8	397	Q92ML7	Q92ML7 rhizobium m
10	38	82.6	925	Q9JG16	Q9JG16 chlamydia p
11	38	82.6	925	Q9ZAP5	Q9ZAP5 chlamydia p
12	38	82.6	1278	Q57236	Q57236 clostridium
13	38	82.6	1280	Q9ZAJ5	Q9ZAJ5 clostridium
14	37	80.4	349	Q8IJV0	Q8IJV0 plasmodium
15	37	80.4	707	Q8X008	Q8X008 neospora
16	36	78.3	105	Q8VXL7	Q8VXL7 fagus sylvia

17	36	78.3	232	16	Q8X829	Q8X829 escherichia
18	36	78.3	237	13	Q8AWC9	Q8AWC9 cyprinus ca
19	36	78.3	241	10	Q8VXL6	Q8VXL6 fagus sylvia
20	36	78.3	489	9	Q9AZH2	Q9AZH2 bacterioph
21	36	78.3	489	16	Q9CEA2	Q9CEA2 lactococcus
22	36	78.3	773	10	Q9SHP2	Q9SHP2 arabidopsis
23	36	78.3	1251	2	Q9KJ95	Q9KJ95 clostridium
24	36	78.3	1252	2	Q8KZM3	Q8KZM3 clostridium
25	36	78.3	1255	2	Q9FAR6	Q9FAR6 clostridium
26	35	76.1	312	16	Q2G068	Q2G068 helicobacte
27	35	76.1	312	16	Q9ZJ59	Q9ZJ59 helicobacte
28	35	76.1	426	5	Q8MYP4	Q8MYP4 caenorhabdi
29	35	76.1	431	5	Q9XTZ9	Q9XTZ9 caenorhabdi
30	35	76.1	500	16	Q9RVQ8	Q9RVQ8 deinococcus
31	35	76.1	679	16	Q97SL8	Q97SL8 streptococc
32	35	76.1	737	16	Q8DR83	Q8DR83 streptococc
33	34	73.9	129	3	Q8TFT7	Q8TFT7 ustilago vi
34	34	73.9	198	2	Q9RCP8	Q9RCP8 streptococc
35	34	73.9	198	2	Q9RCQ5	Q9RCQ5 streptococc
36	34	73.9	198	2	Q9RCR0	Q9RCR0 streptococc
37	34	73.9	198	2	Q9R2M1	Q9R2M1 streptococc
38	34	73.9	198	2	Q9RCQ2	Q9RCQ2 streptococc
39	34	73.9	199	16	Q8DV99	Q8DV99 streptococc
40	34	73.9	222	16	Q8EEB7	Q8EEB7 streptococc
41	34	73.9	225	5	Q8I1V8	Q8I1V8 shewanella
42	34	73.9	261	5	Q8IFQ1	Q8IFQ1 plasmodium
43	34	73.9	299	17	Q96YM4	Q96YM4 sulfolobus
44	34	73.9	385	10	Q8H3N0	Q8H3N0 cryza sativ
45	34	73.9	458	16	Q8FF64	Q8FF64 escherichia

## ALIGNMENTS

RESULT 1

Q9R631 ID Q9R631 PRELIMINARY; PRT; 451 AA.  
AC Q9R631;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TRENBLrel. 23, Last annotation update)  
DE Neurotoxin type B light chain, BONT/B.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92340509; PubMed=1634516;  
RA Kurazono H., Mochida S., Binz T., Eisel U., Quanz M., Grebenstein O.,  
RA Wernars K., Foulain S., Tauc L., Niemann H.;  
RT "Minimal essential domains specifying toxicity of the light chains of  
RL tetanus toxin and botulinum neurotoxin type A.";  
J. Biol. Chem. 267:14721-14729 (1992).  
DR HSSP; P10845; 3BTA.  
DR InterPro; IPR000195; Bontoxilysin.  
DR InterPro; IPR006025; Zn\_MTPeptide.  
DR Pfam; PF01742; Peptidase\_M27; 1.  
DR PRINTS; PR00760; BONTOXILYSIN.  
DR PRODOM; PD001963; Bontoxilysin; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
SQ SEQUENCE 451 AA; 51943 MW; 6C79FD488653EA71 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 451;  
Best Local Similarity 87.5%; Pred. No. 6.3;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLHVH 8  
Db 230 HDLHVH 237

RESULT 2

```

Q45851
ID Q45851 PRELIMINARY; PRT; 1268 AA.
AC Q45851;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type F.
GN BONT /F.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1561;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=9325228; PubMed=8486245;
RA Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA Richardson P.T.;
RT "Nucleotide sequence of the gene coding for Clostridium barati type F
RL neurotoxin: Comparison with other clostridial neurotoxins.";
RL FEMS Microbiol. Lett. 108:175-182 (1993).
DR EMBL; X68262; CAA48329.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_Mtpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1268;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 219 HELIHLVH 226
|:|||||

RESULT 3
Q9ZAJ8 PRELIMINARY; PRT; 1291 AA.
ID Q9ZAJ8;
AC Q9ZAJ8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Bont protein.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CDC 3281;
RA MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318 (1998).
DR EMBL; Y13630; CAA73968.1; -.
DR HSSP; P10845; 3BTA.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_Mtpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150840 MW; E4D3B0E46AB2E735 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 HDLIHVLH 8
Db 230 HELIHLVH 237
|:|||||

RESULT 4
Q93G71 PRELIMINARY; PRT; 1291 AA.
ID Q93G71;
AC Q93G71;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurotoxin type B.
GN Clostridium botulinum.
OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=1436;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF295826; AAK97132.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_Mtpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150824 MW; D7CA07BAE2EB8CD2 CRC64;

Query Match 91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 230 HELIHLVH 237
|:|||||

RESULT 5
Q933K0 PRELIMINARY; PRT; 1291 AA.
ID Q933K0;
AC Q933K0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type B cryptic neurotoxin.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=593, and 588;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF300466; AAL11499.1; -.
DR EMBL; AF300465; AAL11498.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_Mtpeptdse.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150843 MW; 7AC1737B0FA5A151 CRC64;

```



```

Query Match          91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 230 HELIHVLH 237

RESULT 6
Q08077 Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE BONT/B.
DE BONT/B.
GN Clostridium botulinum.
OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eklund 17B ATCC25765;
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT Clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150513 MW; 71BCAEPF23D6FAAA CRC64;

Query Match          91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 230 HELIHVLH 237

RESULT 7
Q08077 Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Neurotoxin.
GN BONTB.
OS Clostridium botulinum.
OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Eklund 17B ATCC25765;
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT Clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1291 AA; 150513 MW; 71BCAEPF23D6FAAA CRC64;

Query Match          91.3%; Score 42; DB 2; Length 1291;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 230 HELIHVLH 237

RESULT 8
Q08077 Q08077 PRELIMINARY; PRT; 1310 AA.
AC Q08077;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Tetanus toxin (Fragment).
GN Clostridium tetani.
OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Shurin Z., Dianliang L.;
RX MEDLINE=94122659; PubMed=7764370;
RA Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
RT Clostridium botulinum type B neurotoxin: comparison with other
RT Clostridial neurotoxins.";
RL Curr. Microbiol. 28:101-110(1994).
DR EMBL; X71343; CAA50482.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase M27; 1.
DR PRINTS; PR00760; BONTXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 1310 AA; 150316 MW; 9EADDC914418E450 CRC64;

Query Match          91.3%; Score 42; DB 2; Length 1310;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HDLIHVLH 8
Db 234 HELIHVLH 241

RESULT 9
Q08077 Q08077 PRELIMINARY; PRT; 397 AA.
AC Q08077;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Putative deaminase OR deamidase protein.
GN R02596 OR SMC02420.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OS Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Bolstad D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Leilaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591791; CAC47175.1; -
DR InterPro; IPR006680; Amidohydro_1.
DR Pfam; PF01979; Amidohydro_1; 1.
DR PROSITE; PS01137; TATD_1; 1.
KW Complete proteome.
SQ SEQUENCE 397 AA; 43054 MW; B7D5F69C499CBE02 CRC64;

Query Match      84.8%; Score 39; DB 16; Length 397;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
Db 326 HDLEHVLH 333

RESULT 10
Q9JS16 PRELIMINARY; PRT; 925 AA.
ID Q9JS16
AC Q9JS16
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE C7234 hypothetical protein.
GN CP0293 OR CP0465.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AE002208; AAF38302.1; -.
DR EMBL; AP002546; BAA98503.1; -.
DR TTGR; CP0465; BAA98503.1; -.
SQ SEQUENCE 925 AA; 105601 MW; 61E8941E7C8FD620 CRC64;

Query Match      82.6%; Score 38; DB 16; Length 925;
Best Local Similarity 62.5%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
Db 538 HDLHLHITH 545

RESULT 11
Q9Z8P5 PRELIMINARY; PRT; 925 AA.
ID Q9Z8P5
AC Q9Z8P5
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE C7234 hypothetical protein.

```

```

GN CPN0293.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
DR EMBL; AE001614; AAD18442.1; -.
KW Complete proteome.
SQ SEQUENCE 925 AA; 105615 MW; 98E6098E7C8FD37D CRC64;

Query Match      82.6%; Score 38; DB 16; Length 925;
Best Local Similarity 62.5%; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HDLIHVLH 8
Db 538 HDLHLHITH 545

RESULT 12
Q57236 PRELIMINARY; PRT; 1278 AA.
ID Q57236
AC Q57236; Q45863;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Botulinum neurotoxin type F (BONT/F protein).
GN BONT/F.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1494;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 10281;
RA Hutson R.A., Collins M.D.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 635-1000 FROM N.A.
RC STRAIN=NCTC 1028;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulin neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN=LANGLEND;
RX MEDLINE=98404102; PubMed=9732534;
RA East A.K., Bhandari M., Hiehl S., Collins M.D.;
RT "Analysis of the botulinum neurotoxin type F gene clusters in
RT proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT barati.";
RL Curr. Microbiol. 37:262-268(1998).
DR EMBL; X81714; CAA57358.1; -.
DR EMBL; L35496; AAA23210.1; -.
DR EMBL; X70821; CAA50152.1; -.
DR EMBL; X99064; CAA67512.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR006025; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase_M27; 1.

```

Search completed: November 12, 2003, 09:36:51  
Job time : 31 secs